Psychometrika

CONTENTS

VOLUME THIRTEEN DECEMBER 1948 NUMBER FOUR

PSYCHOMETRIKA—VOL. 13, NO. 4
DECEMBER, 1948

WAR-MOODS: II

LEWIS F. RICHARDSON
HILLSIDE HOUSE, KILMUN, ARGYLL, BRITAIN

PART II A

Introduction to Part II A

This is a continuation of a paper in the previous issue of *Psychometrika* (Sept., 1948, pp. 147-174) in which some historical facts about the First World War were expressed in mathematical form. The present introduction is intended to make the sequel readable by a person who has never seen Part I. This can perhaps just be done; although for ease, clarity, and conviction a prior study of Part I is strongly to be recommended. The independent variable is t, the time. There are supposed to be two opposing nations, or alliances, distinguished by suffixes 1 and 2, their populations initially numbered N_1 and N_2. The numbers of war-dead at any time, t, are θ_1^*, and θ_2^*. When these are expressed as fractions of their respective populations, then the asterisk is omitted, thus:

$$\theta_1^*/N_1 = \theta_1\,, \quad \theta_2^*/N_2 = \theta_2\,.$$

Similarly, an asterisk applied to the symbol for any fraction of the population makes it into the symbol for the corresponding number of persons. It is characteristic of the present theory, in contradistinction to those of Volterra (41) and Rashevsky (27), that θ_1 and θ_2 are both small fractions. Defeat is here regarded as coming, not by extermination, but by a feeling of hopelessness. The other dependent variables are the fractions of the population in various war-moods. Each mood is regarded as dual, having an overt part, and a concealed, subconscious, or unconscious part. The names of the two parts of a mood are placed in the same bracket, the overt part standing either above or before the other. The whole course of the moods from peace through war to peace again is summarized thus:

$$\begin{array}{ccccccccc} & \text{Arms-race} & & \text{Outbreak} & & \text{Attrition} & & \text{Armistice} & \\ \left.\begin{matrix}\text{Friendly}\\ \text{Friendly}\end{matrix}\right\} & \to & \left\{\begin{matrix}\text{Friendly}\\ \text{Hostile}\end{matrix}\right\} & \to & \left\{\begin{matrix}\text{Hostile}\\ \text{Friendly}\end{matrix}\right\} & \to & \left\{\begin{matrix}\text{Hostile}\\ \text{War-weary}\end{matrix}\right\} & \to & \left\{\begin{matrix}\text{War-weary}\\ \text{Hostile}\end{matrix}\right. \\ \beta & & \xi & & \eta & & \rho & & \omega \end{array}$$

Under each dual mood is placed the Greek letter which represents the fraction of the population in that mood at time t. These letters are to be given suffixes 1 or 2 according to the side to which they belong.

In Part IB the successive changes of mood have been discussed separately. Each process was regarded as a mental infection, after the manner of the simplest of Kermack and McKendrick's (15) theories of epidemics of disease. Each process was provided with an adjustable constant. Of these w is a pure number; but each of A, K, C, E, B, D, F, is the reciprocal of a time. The orders of magnitude of several of them have been estimated with reference to the First World War. If a change of mood is purely internal to a side, then its constant is given a single suffix 1 or 2; if instead the change of mood involves both sides, then its constant is given a double suffix 12 or 21. The meanings of the constants can be gathered from their situation in the following equations.

Because Part II contains several references to Figures 1 and 2, they are reprinted here.

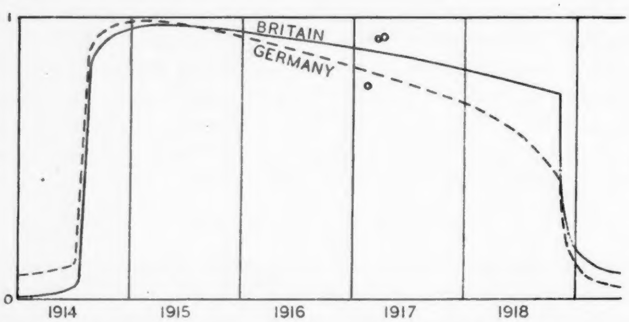
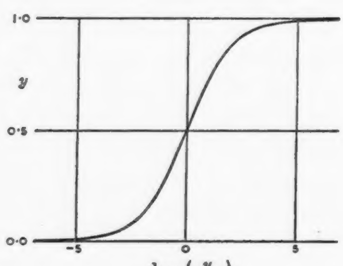


FIG. 1.—The ordinate ζ represents the fraction of the population that was overtly in favor of the war of 1914-18. This diagram is based on historical facts, not on mathematical theory.
Circles mark the British by-elections.



War Fever.

FIG. 2. —Simple approximation to the outbreak of a symmetrical war. The abscissa is proportional to the time. The ordinate is the fraction, of either population, overtly in favor of war.

The complete system of hypotheses

The results of the piecemeal discussion in Part I have now to be collected and fitted together without contradicting one another. Each of the changes is a transfer of persons from one category to another, so that every term must appear twice with opposite signs. Some of the equations of Part I are complete statements and are simply copied here: thus (26) is (7); (28) is (13); (34) is (19); and (38) is (25).* On the contrary, (16) and (23) are partial statements, which are now completed so as to give (30), (32), and (36). The time-rates are printed for the first nation. Those for the second nation can be obtained by interchanging the suffixes, and so are sufficiently represented by merely giving them reference numbers.

$$\beta_1 + \xi_1 + \eta_1 + \theta_1 + \rho_1 + \omega_1 = 1 \qquad (26),(27)$$

$$d\beta_1/dt = \beta_1\{A_1(\xi_1 + \eta_1) - K_{12}(\xi_2 + w_{12}\eta_2)\} \qquad (28),(29)$$

$$d\xi_1/dt = -\beta_1 A_1 \xi_1 + \beta_1 K_{12}(\xi_2 + w_{12}\eta_2) - C_{12}\xi_1(\eta_2 + \rho_2) \qquad (30),(31)$$

$$d\eta_1/dt = -\beta_1 A_1 \eta_1 + C_{12}\xi_1(\eta_2 + \rho_2) - B_{12}\eta_1(\eta_2 + \rho_2) - E_{12}\eta_1(\eta_2 + \rho_2) \qquad (32),(33)$$

$$d\theta_1/dt = E_{12}(\eta_1 + \rho_1)(\eta_2 + \rho_2) \qquad (34),(35)$$

$$d\rho_1/dt = -D_1\rho_1\omega_1 - F_{12}\rho_1\omega_2 + B_{12}\eta_1(\eta_2 + \rho_2) - E_{12}\rho_1(\eta_2 + \rho_2) \qquad (36),(37)$$

$$d\omega_1/dt = D_1\rho_1\omega_1 + F_{12}\rho_1\omega_2 \qquad (38),(39)$$

Here w_{12} is a positive, pure number, regarded as constant. Each capital letter in (26) to (39) is positive, is the reciprocal of a time, and is regarded for simplicity as a constant, except that E_{12} vanishes outside the period of hostilities. This verbal statement will be referred to as - - - - - - - - - - - - - - - (40),(41).
These statements (26) to (41), if found to be true, or even tolerable approximations, will be the most valuable part of the theory; for they express a psychological analysis of observations which first come to our notice in quite other forms. But, before the statements (26) to

*We have so many quantities that some readers may like to have a mnemonic for the symbols. The following is offered on the understanding that it is rather silly and must never be allowed to take the place of the definitions which have already been given: β for beatitude, ξ for excitation, η for eagerness, θ for thanatos, ρ for rueful rumination, and ω because it is the last; A and K because they are respectively related, as will be shown, to the a and k of Richardson's linear theory of arms-races, C because it is like K, E for efficiency in extermination, B for boredom, D for defeatism, F for foreign infection, and w for weight.

(41) can gain credence, they must be subjected to a variety of tests, some simple, some elaborate. Let us begin with the simplest.

Constancy of total population

By adding equations (28), (30), (32), (34), (36), and (38), we obtain

$$\frac{d}{dt}(\beta_1 + \xi_1 + \eta_1 + \theta_1 + \rho_1 + \omega_1) = 0,$$

which agrees with (26).

Buffering

It is necessary to make sure that the hypotheses automatically prevent any one of β_1, ξ_1, η_1, θ_1, ρ_1, ω_1 from over-running its termini 0 and 1. If any one of these six variables is equal to unity we see from (26) that each of the other five must be zero. Several over-runnings are prevented by zero factors arising in this way. To express the results of the test compactly, let x stand for any one of β_1, ξ_1, η_1, θ_1, ρ_1, ω_1. The indications in the body of the following table refer to dx/dt and are deductions from equations (26) to (38).

If x stands for	β_1	ξ_1	η_1	θ_1	ρ_1	ω_1
then at $x = 1$, dx/dt is	$\leqslant 0$	$\leqslant 0$	$\leqslant 0$	$= 0$	$\leqslant 0$	$= 0$
and at $x = 0$, dx/dt is	$= 0$	$\geqslant 0$	$\geqslant 0$	$\geqslant 0$	$\geqslant 0$	$\geqslant 0$

It is seen that when any variable is at one of its termini it is, according to the hypotheses (28) to (39), either stationary or moving into the permitted range. Some, at first sight, harmless changes in the hypotheses would infringe this necessary condition. For example, the author at first supposed that the decrease of β_1 represented by (28) was entirely compensated by an increase of ξ_1. But if so, $d\xi_1/dt$ would contain a term $-\beta_1 A_1 \eta_1$, which could be negative when ξ_1 was zero, contrary to the requirements of buffering. This term had therefore to be placed instead in the equation (32) for $d\eta_1/dt$, where there is no difficulty, as $\beta_1\eta_1$ vanishes when $\eta_1 = 0$ or 1. The remainder of $d\beta_1/dt$ is compensated in $d\xi_1/dt$.

DEDUCTIONS

General remarks on methods of deduction

The hypotheses have now passed the easier tests and are therefore worthy to be tested by more elaborate deductions. The author is

not aware of any explicit general solution of the fourteen equations (26) to (39). We are, however, able to gain considerable insight into the form of integrals by assuming, in accordance with observation, that certain variables are negligible at certain phases in the process. For the majority of the population passes in turn through the moods: {friendly, friendly}, {friendly, hostile}, {hostile, friendly}, {hostile, weary}, {weary, hostile}; so that in turn β, ξ, η, ρ, ω exceed $\frac{1}{2}$; and the other variables are left with less than $\frac{1}{2}$ between them. We can study:

(i) an early phase of an uneasy peace, in which β is near unity but θ, ρ, and ω are zero.

(ii) the outbreak of hostilities, during which ξ and η are important but θ, ρ, and ω are negligible.

(iii) a middle phase of persistence and attrition, in which ξ, η, and ρ are important but β and ω can be neglected.

(iv) the cessation of hostilities at a time when ρ and ω are important but β and ξ can be neglected.

Another device for simplifying the mathematics is to halve the number of differential equations by considering a war between two nations that are equal, at each instant, in all the respects which are represented by symbols in the theory. Such a theoretical scheme will be called for short a *symmetrical war*. It involves no victory-and-defeat, but a process of equal attrition until both nations are simultaneously willing to agree. The first three years of the Great War of 1914-'18 were moderately symmetrical as between France and Britain on the one side and Germany and Austria-Hungary on the other; but marked asymmetry developed at the end. Symmetrical wars are particularly worthy of attention because it will be shown in Case V that, if the constants are symmetrical, the variables tend to become so in the earliest phase, though not necessarily in the final phase.

At first sight some of the more general statements in Volterra's (41) theory of competing species may appear to include the present hypotheses as a special case. Volterra was mainly concerned with populations whose total varies. Nevertheless, constancy can be arranged by putting Volterra's $\beta_i = 1$. There remains the difficulty that in Volterra's most general hypothesis (41, p. 127, eqn. 44), which in his notation reads

$$\frac{dN_i}{dt} = f_i(N_1, \cdots, N_n)N_i \qquad (i = 1, 2, \cdots, n),$$

the N_i appears as a factor in the second member, whereas ξ_1 is not a factor of my equation (30) for $d\xi_1/dt$, nor are η_1, θ_1, ρ_1 and ω_1 factors, respectively, of my equations for $d\eta_1/dt$, $d\theta_1/dt$, $d\rho_1/dt$, $d\omega_1/dt$. Of course, this difficulty could be got over by choosing special forms for f_i; but Volterra appears to exclude such forms by the hypothesis that f_i is continuous when $N_r = 0$. Actually no application has been made of Volterra's formulas.

Case I. The earliest phase

Let us suppose that each nation consists entirely of those who are overtly friendly towards the other nation, but that a small fraction of these people harbor a subconscious hostility. The hypotheses (26) to (35) simplify in this case to

$$\beta_1 + \xi_1 = 1, \qquad \beta_2 + \xi_2 = 1 \tag{42),(43}$$

$$d\beta_1/dt = \beta_1 A_1 \xi_1 - \beta_1 K_{12} \xi_2, \qquad d\beta_2/dt = \beta_2 A_2 \xi_2 - \beta_2 K_{21} \xi_1$$
$$= - d\xi_1/dt \qquad\qquad = - d\xi_2/dt \tag{44),(45}$$

Wherever β_1, β_2 occur as multipliers, we may replace them approximately by unity, - - - - - - - - - - - - - - (46)

so obtaining linear equations

$$d\xi_1/dt = - A_1 \xi_1 + K_{12} \xi_2, \qquad d\xi_2/dt = K_{21} \xi_1 - A_2 \xi_2 . \tag{47) (48}$$

From these it can be proved, as in Richardson (30, §2-11) that ξ_1 and ξ_2 will tend to zero as $t \to \infty$, provided that

$$A_1 A_2 > K_{12} \cdot K_{21} . \tag{49}$$

If (49) were satisfied, peace would be stable. We have, however, seen a reason, namely, the mote and beam effect as expressed in inequality (14), to expect that $K_{12} > A_1$ and $K_{21} > A_2$. If so, ξ_1 and ξ_2 would increase until our approximation (46) became invalid.

Comparison with the linear theory of arms races, as published by Richardson (30 and 32).

To make this comparison we require to express x_1, the psychological attitude of the first nation towards the second, as a function of β_1^* and ξ_1^*. In conformity with Ockham's "razor," let us choose the simplest formulas that make sense, and accordingly put

$$x_1 = - a_1 \beta_1^* + b_1 \xi_1^*, \quad x_2 = - a_2 \beta_2^* + b_2 \xi_2^*, \tag{50),(51}$$

where a_1, b_1, a_2, b_2 are constants. The sign of x_1 or x_2 was conventionally taken throughout the linear theory of arms-races to be

positive if the feeling was hostile, negative if friendly. Accordingly, each of a_1, b_1, a_2, b_2 is positive. - - - - - - - - - - - (52)

Ten million persons, all of the same opinion, would naturally be regarded as having a more intense outward attitude than 100 persons of the same opinion. As a more definite assumption, let us for simplicity suppose that the outward attitude is proportional to the number of persons of like mind - - - - - - - - - - (53)

Now β_1^*, ξ_1^*, β_2^*, ξ_2^* are numbers of persons. So in (50) and (51) the constants a_1, b_1, a_2, b_2 do not depend on how many people there are. - - - - - - - - - - - - - - - - - - - (54)

They may be called personal as distinct from national constants.

On transformation to fractions of the population, (50) and (51) become

$$x_1/N_1 = -a_1\beta_1 + b_1\xi_1\,, \quad x_2/N_2 = -a_2\beta_2 + b_2\xi_2\,. \quad (55),(56)$$

Let β_1, β_2 be eliminated by means of (42) and (43), so that

$$x_1/N_1 = -a_1 + (a_1+b_1)\,\xi_1\,, \quad x_2/N_2 = -a_2 + (a_2+b_2)\,\xi_2\,. \quad (57),(58)$$

From now on it will suffice to attend to the first of each pair of equations. The second, which can be obtained by interchanging the suffixes, is given a reference number to remind us that it is available.

The derivative of (57) is

$$\frac{d(x_1/N_1)}{dt} = (a_1 + b_1)\,\frac{d\xi_1}{dt}. \qquad (59),(60)$$

Let $d\xi_1/dt$ be eliminated between (59) and (47), so that

$$\frac{d(x_1/N_1)}{dt} = (a_1 + b_1)\,(-A_1\xi_1 + K_{12}\xi_2)\,. \qquad (61),(62)$$

Finally, in (61) let ξ_1 and ξ_2 be expressed in terms of x_1 and x_2 by way of (57) and (58), giving

$$\frac{d(x_1/N_1)}{dt} = (a_1 + b_1)\left\{-\frac{A_1}{a_1 + b_1}\left(\frac{x_1}{N_1} + a_1\right) + \frac{K_{12}}{a_2 + b_2}\left(\frac{x_2}{N_2} + a_2\right)\right\}. \qquad (63),(64)$$

This is a linear equation comparable with the fundamental assumption of Richardson's (30, p. 6) linear theory of arms-races, namely,

$$dx_1/dt = -\alpha_1 x_1 + k_{12}x_2 + g_1\,. \qquad (65),(66)$$

On comparison of the coefficients of the variables in (63) with those in (65), we see that:

The fatigue-and-expense coefficient, α_1, is given by $\alpha_1 = A_1$. (67),(68)

The defense-coefficient k_{12} is given by

$$k_{12} = K_{12}\left(\frac{a_1 + b_1}{a_2 + b_2}\right)\frac{N_1}{N_2}. \qquad (69),(70)$$

The "grievance," g_1, is given by

$$g_1 = N_1\left\{-A_1a_1 + K_{12}\left(\frac{a_1 + b_1}{a_2 + b_2}\right)a_2\right\}. \qquad (71),(72)$$

These formulas are of considerable interest. A new psychological interpretation is given to α_1 and to A_1 by their equality. Formerly α_1 was supposed to depend on the fatigue of personal service in preparation for war and on a reluctance to incur the expense of armaments; whereas A_1 was supposed to be an expression of persistence in friendliness in spite of threats. It will be better in future to regard α_1 and A_1 as expressing the joint effect of all such restraining influences and to call them *"restraint coefficients."* This statement will be referred to as - - - - - - - - - - - - - - - - (73).
If we eliminate A_1 and K_{12} between (67), (69), and (71), we obtain

$$g_1 = -N_1a_1\alpha_1 + N_2a_2k_{12}. \qquad (74),(75)$$

Thus g_1 is an independent constant, not expressible in terms of α_1 and k_{12} alone. Also, g_1 is capable of either sign. Both these verbal results accord with the linear theory of arms-races. But the relation (74) makes also other assertions, which are new, and problematic.
On multiplying (69) by (70), we have

$$k_{12} \cdot k_{21} = K_{12} \cdot K_{21}. \qquad (76)$$

So in view of (67) and (68),

$$k_{12} \cdot k_{21}/\alpha_1\alpha_2 = K_{12} \cdot K_{21}/A_1A_2, \qquad (77)$$

that is to say, the conditions of stability in the two theories are consistent with one another.

The orders of magnitude of A_1 *and* K_{12} can be estimated by way of (67) and (69).

The numerical value of α was shown by Richardson (32), from the rate of demobilization, to be about one year^{-1}. Therefore:

The constant A_1 is of the order of one year^{-1}. - - - (78)

Of the constants that occur in (69), we may expect from the definition in (50) and (51) that:

a_1 is of the order of a_2, and b_1 of the order of b_2, so that $(a_1 + b_1)/(a_2 + b_2)$ is of the order of unity. - - - - (79)

In picking out a defense coefficient from the observations in Richardson (32), for comparison with K_{12}, we must pay attention to the note at the end of Type IX, which is to the effect that it may be proper to regard the world as composed either of two alliances as was done in one section, or of ten nations as was done in another; but that the defense coefficients appropriate to these two views will be different. So it is perhaps clearest to state simply that: if the nations of the pair are equal, so that $N_1 = N_2$, and if their instability coefficient is λ, then

$$k_{12} = \alpha + \lambda\text{; and so by (67) and (69) } K_{12} = A + \lambda\,, \tag{80}$$

where λ has been of the order of $0._3$ year^{-1} during the two great arms-races.

When the nations are very unequal we should attend to K_{21}/K_{12}. By dividing (69) by (70), we obtain

$$\frac{k_{12}}{k_{21}} = \frac{K_{12}}{K_{21}}\left(\frac{a_1 + b_1}{a_2 + b_2}\right)^2\left(\frac{N_1}{N_2}\right)^2. \tag{81}$$

But reasons were given by Richardson (30, pp. 74, 75, revised in 32) for believing the defense coefficient to be proportional to the "size" of a nation in the sense of its industry and population. The same idea appeared in the microfilm (32) in Type IX and Type XII, and was widely applied. As far as it can be expressed in terms of population alone, it is that

$$k_{12}/k_{21} = N_1/N_2\,, \tag{82}$$

the suffix of N corresponding to the *first* suffix of k. When this is introduced into the previous equation, there is left

$$\frac{N_2}{N_1} = \frac{K_{12}}{K_{21}}\left(\frac{a_1 + b_1}{a_2 + b_2}\right)^2, \tag{83}$$

in which the suffix of N corresponds to the *second* suffix of K.

In view of (79), we conclude that

$\mathrm{K}_{21}/\mathrm{K}_{12}$ *is probably of the order of* $\mathrm{N}_1/\mathrm{N}_2$. (84)

Compare statement (22) on page 171.

Case II. Simple approximation to the outbreak of a symmetrical war

As the two nations are assumed to be equal in respect of each quantity that occurs in the theory, we may omit the suffixes. For simplicity, it will be supposed that there has been a long arms-race in the course of which nearly everybody has changed from the mood {friendly, friendly} to the mood {friendly, hostile}; so that when the present phase begins β is negligible and ξ is near to unity. Near the beginning of a war ω is zero and ρ and θ not far from it. Observation shows that the processes of attrition, represented by $d\theta/dt$ and $d\rho/dt$ are very slow in comparison with the sudden uprising of indignation represented by $d\eta/dt$. We may therefore neglect B and E in comparison with C. The hypotheses (26) to (41) accordingly reduce to

$$\xi + \eta = 1 \tag{85}$$

$$d\xi/dt = -C\xi\eta = -d\eta/dt. \tag{86}$$

On eliminating ξ between (85) and (86) and separating the variables, we obtain

$$Ct + \text{Const.} = \int \frac{d\eta}{\eta(1-\eta)} = \log_e\left(\frac{\eta}{1-\eta}\right). \tag{87}$$

A graph of this equation is shown in Fig. 2 (on page 198) for the special case $C = 1$, const. $= 0$, by a graph of y as a function of $\log_e\{y/(1-y)\}$. The graph has a shape suitable for representing the observations insofar as: (i) the straight line $\eta = 0$ is an asymptote, (ii) from which the curve rises continuously, (iii) with the fewest possible changes of sign of its curvature, (iv) to approach the straight line $\eta = 1$ as another asymptote, and (v) with any degree of suddenness, or of its opposite, obtainable by adjustment of the constant C. There are plenty of other functions which are also suitable because they possess the properties (i) to (v). The rather vague observational data are insufficient to decide the choice between these functions. We may accept (87) on this understanding, noting also that, when θ and ρ are taken into account, η will have a maximum near unity instead of tending towards an asymptote. A maximum is shown in Case VI.

In order to determine the value of C, let $\eta = \eta'$ when $t = t'$, and $\eta = \eta''$ when $t = t''$. Then from (87)

$$C = \frac{1}{t'' - t'} \log_e \frac{\eta''(1-\eta')}{\eta'(1-\eta'')}. \tag{88}$$

With reference to Figure 1 and the collection of historical facts in Part IA, we may reasonably assume that between July 23, 1914 and August 4, 1914, there was in Britain a rise from $\eta' = 0.1$ to $\eta'' = 0.9$, which occurred in the course of about 10 days; so that $t'' - t' = 0.028$ year. When these particulars are inserted in (88) they give

$$C = 1.6 \times 10^2 \text{ year}^{-1}. \tag{89}$$

On referring back to equations (78) and (80) we notice that C, which expresses a rate of conscious change, is of the order of 100 times K, which expresses a rate of subconscious change.

Case III. The phase of attrition in the middle of a symmetrical war

Let us suppose that there are no people left in the overtly friendly moods; so that $\beta = 0$ and $\xi = 0$; but that on the other hand no one yet admits a wish to make peace; so that $\omega = 0$. With these assumptions the hypotheses (26) to (41) reduce to

$$\eta + \theta + \rho = 1 \tag{90}$$

$$d\eta/dt = -(B + E)\eta(\eta + \rho) \tag{91}$$

$$d\theta/dt = E(\eta + \rho)^2 \tag{92}$$

$$d\rho/dt = (B\eta - E\rho)(\eta + \rho). \tag{93}$$

Let $\eta + \rho$ be eliminated between (90) and (92), giving

$$d\theta/dt = E(1 - \theta)^2, \tag{94}$$

the integral of which is

$$Et + \text{Const.} = 1/(1 - \theta). \tag{95}$$

Let $\theta = \theta'$ at $t = t'$ and $\theta = \theta''$ at $t = t''$. It follows, on elimination of the arbitrary constant, that

$$E = \frac{\theta'' - \theta'}{(t'' - t')(1 - \theta')(1 - \theta'')}. \tag{96}$$

This agrees with the rougher estimate of equation (20) to within a few per cent.

For example, the Italian war-deaths among both the military and the civil population are given for a period in the middle of the Great War by Vedel-Petersen (39, pp. 144, 165). At the beginning of 1916, $\theta^* = 1.0 \times 10^5$, if one distributes the deaths of prisoners. Two years later, $\theta^* = 4.6 \times 10^5$. Meanwhile $N_1 = 36.7 \times 10^6$; so that $\theta' = 0.0027$, $\theta'' = 0.0125$. When these data are inserted in (96), it gives

$$E = 0.005 \text{ year}^{-1}. \tag{97}$$

The separate casualties for 1916 and 1917 show that E varied in a ratio of 1.3 between those years.

Let $t = 0$ at the beginning of hostilities when both $\theta = 0$ and $\rho = 0$. In view of (90), we have also

$$\eta = 1 \text{ at } t = 0. \tag{98}$$

It may seem unnatural to suppose that $\eta = 1$, for a nation is seldom quite unanimously in favor of war. But we can avoid that difficulty by supposing that the instant $t = 0$, though used as a reference mark, may lie outside the phase for which the approximations of the present case are valid; so that the formulas are to be applied only for values of t that are not too near to zero. With this origin of time, the arbitrary constant in (95) is unity, so that

$$1 - \theta = 1/(Et + 1). \tag{99}$$

By (90) and (99), $\eta + \rho = 1/(Et + 1)$, and when this is substituted in (91), the variables can be separated, giving, with the convention (98),

$$\log \eta = -\frac{B + E}{E} \log (Et + 1). \tag{100}$$

From (99) and (100),

$$\eta = (1 - \theta)^{1+B/E}. \tag{101}$$

This formula appears to be suitable for representing the facts insofar as it shows that η will steadily diminish at a rate which is controlled by the constant B. The difficulty in the way of determining B from observations is that the persons counted in ρ^* differ qualitatively from those counted in η^* only by a change in the subconscious, which they usually conceal, and concerning which therefore no statistics were, or could have been, collected at the time of its occurrence. From admissions made at a later phase, say just after the end of the Great War, it may be gathered, rather vaguely, that a majority of the population had passed through the mood {aggressive, war-weary}, and so ρ had approached unity and η had approached zero. We can make a conditional estimate of B/E from (101) by supposing that when $\theta = 0.02$, a value typical of several nations towards the end of the war of 1914-'18, then η was, say, η_T. On solving (100), we obtain

$$\frac{B}{E} = \frac{\log \eta}{\log(1-\theta)} - 1 = \frac{\log_{10} \eta_T}{-0.00877} - 1. \tag{102}$$

This relation is shown by pairs of values in the following table:

η_T	0.3	0.1	0.03	0.01
B/E	59	113	173	227

As E, for severely smitten nations, is known from (21) and (97) to be of the order of 10^{-2} year^{-1}, we may conclude that:

$$\text{B } \textit{is of the order of } 1 \text{ year}^{-1} . \tag{103}$$

Case IV. A simple approximation to the final phase of a symmetrical war

Let us suppose that the initial population of N now consists only of $\theta_T{}^{\bullet}$ dead, of $\rho^{\bullet}$ in the mood {aggressive, war-wearied} and of $\omega^{\bullet}$ in the mood {war-wearied, aggressive}; so that

$$\theta_T + \rho + \omega = 1 . \tag{104}$$

Observation shows that the collapse at the end of the First World War, though not as sudden as the uprush of indignation at the beginning, was more rapid than the attrition in the middle phase; so that we may regard θ_T as constant during the collapse. This is equivalent to neglecting E. Accordingly, the hypotheses (28) to (38) reduce to

$$d\rho/dt = -(D+F)\rho\omega , \tag{105}$$

$$d\omega/dt = (D+F)\rho\omega . \tag{106}$$

On eliminating ρ between (104) and (106), we have

$$d\omega/dt = (D+F)(1-\theta_T-\omega)\omega . \tag{107}$$

By separating the variables, we obtain the integral

$$(D+F)t + \text{Const.} = \frac{1}{1-\theta_T}\log\left\{\frac{\omega}{1-\theta_T-\omega}\right\} . \tag{108}$$

We may reasonably suppose that hostilities will come to an end when about half the survivors are unwilling to continue the struggle. To found a precise definition on this rough notion,

$$\text{let } t = T \text{ when } \omega = \tfrac{1}{2}(1-\theta_T) \tag{109}$$

so that $t = T$ at, or about, the end of hostilities. On substituting (109) into (108), we find that the arbitrary constant is $-(D+F)T$, and so

$$(1-\theta_T)(D+F)(t-T)=\log\left\{\frac{\omega}{1-\theta_T-\omega}\right\}. \tag{110}$$

This equation is closely similar to (87); and the same graph, Fig 2, on page 198, can, by suitable interpretation of its scales, be made to represent both. Time runs the same way in either case; but the ordinate represents those openly in favor of war at the outbreak, and those openly in favor of peace at the cessation of hostilities.

Otherwise if $t=t'$ when $\omega=\omega'$ and $t=t''$ when $\omega=\omega''$, then

$$D+F=\frac{1}{(t''-t')(1-\theta_T)}\log_e\left\{\frac{\omega''(1-\theta_T-\omega')}{(1-\theta_T-\omega'')\omega'}\right\}. \tag{111}$$

To determine the order of magnitude of $D+F$ from (111) taken in conjunction with the observational data shown in Fig. 1, we should first draw a curve midway between the curves for Britain and Germany so as to represent an artificially symmetrized war. On this mean curve the ordinate, which is in general $\eta+\rho$, but in this phase simply ρ, falls from 0.67 to 0.15 during the last half of the year 1918. Also, the mean value of θ_T for Britain and Germany was 0.02. These values, when inserted into (104) and (111), yield

$$D+F=5\,\text{year}^{-1}. \tag{112}$$

According to (110), $\omega\to 0$ as $t\to-\infty$. The hypotheses thus entail that ω grows, so to speak, from seed; so that unless there were a few defeatists at the beginning of the war there could not be any at the end.

PATRIOT: "Let us get rid, then, of these accursed pre-war defeatists!"

AUTHOR: "Before you do that, pray let me point out that, if there were no defeatists, the end of this symmetrical war could not come until every person in both nations had been killed.

It may well be a defect of the hypotheses that they have this entailment. The question is not easily decided by observation; for a great variety of small minorities exist. One is reminded of the difficulty that Pasteur had in proving the nonexistence of spontaneous generation of life.

PATRIOT: "May be. But I wasn't thinking of symmetrical wars."

AUTHOR: "When gregariousness is taken into account, as in Case IX, the formulas are different."

Case V. Symmetry in the constants but not necessarily in the variables

It was for mathematical convenience that we considered a perfectly symmetrical system. Actual wars, on the contrary, have mostly ended unsymmetrically in victory and defeat. So let us enquire whether the mathematical model has any corresponding tendency. That is to say, if the two nations are equal in the constants, does an inequality in the variables grow or diminish as time goes on? This question will be answered for two phases separately.

In the *earliest phase,* for which, as in Case I,

$$\beta_1 + \xi_1 = 1\,, \qquad \beta_2 + \xi_2 = 1\,, \qquad (113),(114)$$

the hypotheses (28) to (38) become, when we omit the suffixes on the constants,

$$d \log \beta_1/\, dt = A\xi_1 - K\xi_2\,, \qquad d \log \beta_2/dt = A\xi_2 - K\xi_1\,, \quad (115),(116)$$

whence by subtraction

$$\frac{d \log(\beta_1/\beta_2)}{dt} = (A + K)\,(\beta_2 - \beta_1)\,. \qquad (117)$$

If at any instant $\beta_2 > \beta_1$, then $\log \beta_2 > \log \beta_1$, and so $\log(\beta_1/\beta_2) < 0$. But $A + K$ is positive. So $\log(\beta_1/\beta_2)$ is increasing and is moving towards zero. But if, on the contrary, $\beta_2 < \beta_1$ at any instant, then $\log(\beta_1/\beta_2)$ is positive and is decreasing. Thus (117) entails that $\log(\beta_1/\beta_2)$ is either zero or is always moving towards zero; so that β_1/β_2 is either unity or is moving towards unity. We may express this result by saying that *in the earliest phase symmetry in the variables is stable.* - - - - - - - - - - - - - - - - (118)
Common observation supports this deduction.

Solving for A. Because attempts to determine the restraint-constant α from a study of defense-expenditure have so far failed except for the phase of demobilization (see the microfilm, 32), any hope of determining the theoretically equal quantity A is all the more welcome. On eliminating K between (115) and (116), we obtain

$$A = \frac{\xi_1\, d \log \beta_1/dt - \xi_2\, d \log \beta_2/dt}{\xi_1^2 - \xi_2^2}. \qquad (119)$$

From this we can remove β_1 and β_2 by means of (113) and (14) for

$$\xi_1\, d \log \beta_1/dt = \left[\frac{-\xi_1}{1-\xi_1}\right]\frac{d\,\xi_1}{dt} = -\frac{1}{2(1-\xi_1)}\,\frac{d\,\xi_1^2}{dt}. \qquad (120)$$

In the earliest phase when ξ_1 and ξ_2 are both near to zero, we may replace the factors $1 - \xi_1$ and $1 - \xi_2$ by unity, so that

$$A = -\frac{1}{2}\frac{d}{dt}\log|\xi_1^2 - \xi_2^2|, \text{ approximately.} \tag{121}$$

It has already been shown that $\xi_1^2 - \xi_2^2$ drifts towards zero. If some observations of this drift towards symmetry are found, then (121) or more accurately (119), can be used to determine A.

In the *final phase*, for which θ is treated as in Case IV as a common constant, θ_T, we have

$$\rho_1 + \omega_1 = 1 - \theta_T, \qquad \rho_2 + \omega_2 = 1 - \theta_T. \qquad (122),(123)$$

Also the hypotheses (28) to (38) become

$$d\log\rho_1/dt = -F\omega_2 - D\omega_1, \tag{124}$$

$$d\log\rho_2/dt = -F\omega_1 - D\omega_2, \tag{125}$$

whence by subtraction

$$\frac{d\log(\rho_1/\rho_2)}{dt} = (F - D)(\omega_1 - \omega_2) = (F - D)(\rho_2 - \rho_1). \tag{126}$$

If $F > D$ the equation (126) resembles (117); and arguments similar to those stated above show that ρ_1/ρ_2 moves towards unity.

But if $F < D$, then $\log(\rho_1/\rho_2)$ moves away from zero on either side; and ρ_1/ρ_2 moves away from unity.

That is to say: *in the final phase symmetry in the variables is stable if* F > D, *and unstable if* $F < D$. - - - - - - - (127)

When we look back at the psychological meanings of D and F, we see that they both express the infectiousness of an open admission of war-weariness; but that F relates to an infection coming from the enemy, presumably by radio; whereas D relates to an infection passing between persons inside the same nation by radio, newspaper, and conversation. To the author, the above relation (127) seems credible but new.

An equation like (126) occurred in the theory of the *reciprocal inhibition* between a pair of osglim lamps wired in parallel, reference (29). This opportunity is taken to correct, with apologies, two mistakes in that paper. In the last member of its equation (6), for $+ u_2$, read $- u_2$. In the second line below equation (7), for "second member," read "last bracket."

Case VI. The outbreak of a symmetrical war and the beginning of suppressed weariness

The treatment of the outbreak in Case II was made as easy as possible; but the approximations failed at the time when war-fever, as represented by η, should have reached its maximum. Let us now improve the method by retaining ρ so that the formulas of Case VI may replace those of Case II for the outbreak, and may extend to the phase in which suppressed weariness begins. We may still, however, neglect ω, for it belongs to the last phase of a war. Also, in view of the orders of magnitude of B and E, as found in Case III, we may neglect E while retaining B. Thus the hypotheses (26) to (39) reduce to

$$\xi + \eta + \rho = 1 \tag{128}$$

$$d\xi/dt = -\xi(\eta + \rho)C \tag{129}$$

$$d\eta/dt = \xi(\eta + \rho)C - \eta(\eta + \rho)B \tag{130}$$

$$d\rho/dt = \eta(\eta + \rho)B\,. \tag{131}$$

By (129) $d\xi/dt$ has always the same sign, thus permitting ξ to be used as a parameter defining the time. On elimination of $\eta + \rho$ between (128) and (129), we have

$$-Ct + \text{Const.} = \int \frac{d\,\xi}{\xi(1-\xi)} = \log \frac{\xi}{1-\xi}. \tag{132}$$

A graph of y as a function of $\log\{y/(1-y)\}$ appears in Fig. 2 on page 198, and if read from right to left to suit the negative coefficient of t in (132), that graph shows the manner in which ξ falls from 1 to 0 as t increases from $-\infty$ to ∞. According to (132) and (128), $\eta + \rho \to 1$ as $t \to \infty$. This is a fault of the approximation, arising from the neglect of θ and ω. But it merely indicates that the formulas of the present Case VI must not be applied to the end of a war. Until the maximum of η is passed, the formulas may be expected to be useful approximate deductions from the complete hypotheses.

It is convenient to let $t = 0$ when $\xi = \frac{1}{2}$, - - - - - (133)
for with that choice the arbitrary constant in (132) vanishes, leaving simply

$$t = \frac{1}{C} \log \frac{1-\xi}{\xi}, \tag{134}$$

which is shown in Fig. 3.

Elimination of dt between (129) and (130) gives

$$\frac{d\eta}{d\xi} - \frac{\eta}{\xi}\frac{B}{C} = -1. \tag{135}$$

To determine η as a function of t, we can first integrate (135) so as to express η in terms of ξ, and afterwards bring in the time by way of (134). The integrating factor is $\xi^{-B/C}$; and so the solution of (135) is

$$\eta = \xi^{B/C}\left\{\text{Const.} - \frac{C}{C-B}\,\xi^{(C-B)/C}\right\}. \tag{136}$$

To determine the arbitrary constant in (136), let us assume an initial condition in which at a late stage of the arms race

$$\xi = 1, \qquad \eta = 0. \tag{137}$$

It follows that

$$\eta = \frac{C}{C-B}\,(\xi^{B/C} - \xi), \tag{138}$$

which is also shown in Fig. 3.
From (138)

$$\frac{d\eta}{dt} = \frac{C}{C-B}\left(\xi^{(B-C)/C}\cdot\frac{B}{C} - 1\right)\frac{d\xi}{dt}, \tag{139}$$

and, as $d\xi/dt$ does not vanish for finite values of t, the maximum η_M of η in time occurs when $\xi = \xi_M$, say, such that

$$\xi_M = \left(\frac{C}{B}\right)^{C/(B-C)}; \tag{140}$$

and from (138) and (140)

$$\eta_M = \frac{C}{C-B}\left\{\left(\frac{C}{B}\right)^{B/(B-C)} - \left(\frac{C}{B}\right)^{C/(B-C)}\right\} = \left(\frac{C}{B}\right)^{B/(B-C)}. \tag{141}$$

As a check we have, directly from (135),

$$\eta_M/\xi_M = C/B. \tag{142}$$

From (141) the following pairs of values have been computed

$$\begin{array}{lllll} C/B = 50 & 100 & 200 & 400 & \\ \eta_M = 0.923 & 0.955 & 0.974 & 0.985. & \end{array} \tag{143}$$

In any voting on the question whether a war should be continued, (see Part IA) it is probably not η but $\eta + \rho$ that can be observed, or, in a secret ballot, perhaps some number between η and $\eta + \rho$. Nevertheless we are not entirely ignorant of the actual historical value of

η_M. For η_M must be less than or equal to the greatest fraction of the population ever observed to be in favor of the war. Also observation indicates that there is a time, soon after the outbreak of a war, when enthusiasm for the war is at its greatest and there is no cause to suspect the existence of much suppressed war-weariness, so that at that time $\rho = \rho_M$, say, must be almost zero and η_M roughly observable. This is an interim approximation, for we shall find out more about ρ_M presently. On looking at the graph (Fig. 1) of the course of $\eta + \rho$ in Germany and Britain during the war of 1914-'18 and taking the average for the two countries in order to compare it with the present theory of a symmetrical war, we see that the maximum of $\eta + \rho$ was 0.98 and so $\eta_M \leqslant 0.98$. On comparison with (143), it follows that

$$C/B \leqslant 300\,. \tag{144}$$

But, from (89), $C = 1.6 \times 10^2\ \text{year}^{-1}$.
Therefore

$$B \geqslant 0.5\ \text{year}^{-1}. \tag{145}$$

This estimate is of the same order as that recorded in (103) although obtained by an entirely different method.

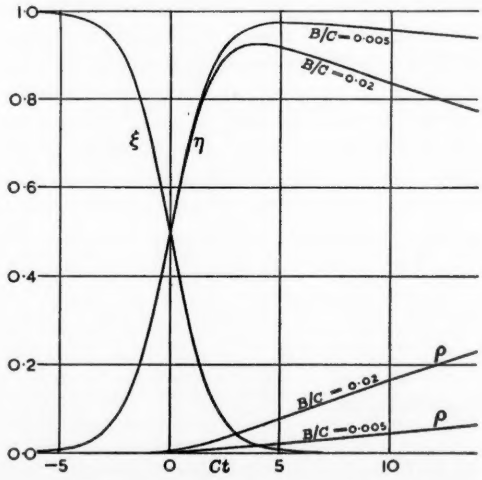


FIG. 3. Outbreak of a symmetrical war and the beginning of weariness, as deduced from hypotheses. Time runs horizontally to the right. The ordinates are fractions of the population. Three moods are shown: ξ marks those who are outwardly friendly but subconsciously hostile: η marks those who are outwardly hostile but subconsciously friendly; ρ marks those who are outwardly hostile but secretly weary. For η and ρ there are two curves, depending on the value of B/C.

For the purpose of drawing graphs to illustrate the formulas let us choose, in conformity with (144),

$$B/C = 0.005 . \tag{146}$$

Then

$$\eta = \frac{1}{0.995} (\xi^{0.005} - \xi) . \tag{147}$$

In Fig. 3 the curves show ξ, η, ρ as functions of t. They have been taken from formulas (134), (138), and (128) for $B/C = 0.005$ and again for $B/C = 0.02$. The graphs of η in particular are suitable for comparison with observations of the fraction of a population that is keen on a war. These graphs are seen to be an improvement on that of Case II because now the curve rises to a maximum and falls slowly. The rate of fall relative to the rate of rise depends on B/C. The time-scale depends on C. If C is given the value in (89), namely, $C = 1.6 \times 10^2$ year^{-1}, then the whole extent of the diagram, which is 20 units of Ct, amounts to $\frac{1}{8}$ year $= 6\frac{1}{2}$ weeks.

A war that is extinguished at its beginning

The maximum attained by η is controlled in the mathematics chiefly by the constant B, which regulates the rate at which the number of secretly weary persons increases. If we make B large enough, η_M can be made as small as we please. It may be possible in this way to represent a mutual flaring up of indignation in two countries such as the Fashoda incident between Britain and France in the autumn of 1898, when war seemed likely, but was avoided. It is not possible, however, to explore this suggestion thoroughly without taking account of ω, the number of those who admit their weariness; for if $\eta_M < 1/2$, then ρ becomes important by the time that η attains its maximum and so ω develops early. As much as is worth doing, in the present neglect of ω, is to note, as a first approximation, that formula (141) shows that $\eta_M = \frac{1}{2}$ if $B/C = \frac{1}{2}$. An alternative explanation of how a war may be just avoided is given in Part II B with reference to gregariousness.

Case VII. An outbreak between unequal nations

As in Case II let us neglect β, ρ, θ, ω; but let us restore from (26), (27), (30), and (31) the suffixes which were not needed in the symmetrical case. The hypotheses now are

$$\xi_1 + \eta_1 = 1 , \qquad \xi_2 + \eta_2 = 1 , \tag{148) (149}$$

$$d\xi_1/dt = -C_{12}\xi_1\eta_2\,, \qquad d\xi_2/dt = -C_{21}\xi_2\eta_1\,. \tag{150) (151}$$

On eliminating dt, we have

$$d\xi_1/d\xi_2 = C_{12}\xi_1\eta_2/(C_{21}\xi_2\eta_1)\,. \tag{152}$$

The variables can be separated, giving

$$C_{21}(1-\xi_1)\, d\log\xi_1 = C_{12}(1-\xi_2)\, d\log\xi_2\,. \tag{153}$$

Whence on integration

$$C_{21}(\log\xi_1 - \xi_1) = C_{12}(\log\xi_2 - \xi_2) + \text{constant}\,. \tag{154}$$

The constant of integration can be fixed by the *supposition that the two nations began the quarrel at the same time in the sense that there was some early time when*

$$\xi_1 = 1 = \xi_2\,. \tag{156}$$

At that instant $-C_{21} = -C_{12} + \text{Constant.}$

Therefore at all times

$$C_{21}(\log\xi_1 - \xi_1 + 1) = C_{12}(\log\xi_2 - \xi_2 + 1)\,. \tag{157}$$

The course of the function is illustrated in the following table.

ξ	1.0	0.9	0.8	0.7	0.6	0.5	0.4
$\log\xi - \xi + 1$	0	—.0054	—.0231	—.0567	—.1108	—.1931	—.3163
ξ	0.3	0.2	0.1	0.05	0.03	0.01	0
$\log\xi - \xi + 1$	—.5040	—.8094	—1.403	—2.046	—2.537	—3.615	$-\infty$

Let this theory now be compared with historical facts concerning some very unsymmetrical outbreaks. The facts are usually given us as incidents separated by pauses. It is reasonable to suppose that the changes of mood were less jerky, and so more like the smooth curves of the theory.

Paul Kruger, President of the South African Republic, records in his *Memoirs* (17) that in the autumn of 1899 prior to his ultimatum of October 9:

> On the 22nd September, the mobilization of an army corps for South Africa was announced in England, and, on the 28th of September, it was announced that the greater part of that army corps would leave for South Africa without delay. The Government thereupon commandeered the greater part of the burghers to take up their position near the frontiers of the Republic, in order to be prepared for a sudden attack on the part of England.

That is to say, the mobilization went on at about the same rate in men on both sides, but very much faster among the Boers when reckoned as a fraction of the population. Mobilization is of course not identical with the change from the mood {friendly, aggressive} to the mood {aggressive, friendly}. Nevertheless we may safely conclude that η_2 among the Boers rose much more rapidly than η_1 among the British.

There is a similar fact about the Russo-Finnish war of 1939-'40. On December 1, 1939 the Russions invaded Finland. But six weeks earlier it was reported that Finland's mobilization was completed (*Glasgow Herald*, October 20, 1939). On November 15th, a fortnight before the invasion, the same paper reported that "One-fifth of Finland's population of 3,000,000 is now under arms." There is no suggestion that a fifth of the Russian population was under arms; and so huge a mobilization could not have escaped notice, had it existed.

It is understandable that in any dispute between very unequal populations, the less numerous people are likely to mobilize first, because they feel more threatened. This is one of the many difficulties in the way of giving any equitable answer to the question: who began it?

Now let us return to equation (157). If at any time $\eta_2 > \eta_1$, then $\xi_2 < \xi_1$ and $|\log \xi_2 - \xi_2 + 1| > |\log \xi_1 - \xi_1 + 1|$; whence $C_{21} > C_{12}$. So the comparison of the Boer-British and Finn-Russian wars with the theory leads to the conclusion that

$$\text{if } N_1 >> N_2 \quad \text{then} \quad C_{21} > C_{12}. \tag{158}$$

To estimate C_{21}/C_{12}, let us consider the time in 1899 when the British people were equally divided between overt friendliness and overt hostility; so that ξ_1 was 0.5 and $\log \xi_1 - \xi_1 + 1$ was -0.1931. According to (22) E_{21}/E_{12} was of the order of $(N_1/N_2)^{0.7}$; and, according to (84), K_{21}/K_{12} was of the order N_1/N_2. So let us now suppose for the sake of argument that $C_{21}/C_{12} = N_1/N_2 = 180$. Then $\log \xi_2 - \xi_2 + 1 = -0.1931 \times 180 = -34.76$ and therefore $\xi_2 = 10^{-15}$. But that is an incredible degree of unanimity; for Botha opposed the Boer ultimation to Britain (7, 3, p. 948). So let us suppose instead that $C_{21}/C_{12} = \sqrt{(N_1/N_2)} = \sqrt{180} = 13.4$. Then $\log \xi_2 - \xi_2 + 1 = -0.1931 \times 13.4 = -2.59$; whence $\xi_2 = 0.029$; which indicates that only three Boers in a hundred were still overtly friendly. This is more credible.

Conclusion.

$$C_{21}/C_{12} \textit{ may be of the order of } \sqrt{(N_1/N_2)}. \tag{159}$$

Summary on the bipersonal theory of war-fever

Practical politics has been called the "art of the possible"; and the same may be said of mathematical politics. This bipersonal theory has been arranged so as to have one constant for each important psychological effect and so that the set of equations is sufficiently tractable. The gist of the theory is contained in the hypotheses (26) to (41). They have been tested by making from them a variety of deductions and by comparing these with the war of 1914-'18. The deductions have been found to be in interesting agreement with the observations. The orders of magnitude of most of the constants have been estimated and are collected in the following table. Support is thus given to the psychological discussion which led up to the formulation of the hypotheses (26) to (41). But only moderate determinism should be expected. For the observations are not definite or consistent enough to confirm or to refute any theory with the precision expected in physics. And the theory can only express what is habitual, traditional, or instinctive; whereas we know, in everyday life, the operation of free choice.

Collection of Constants. Those without suffixes are mostly for a symmetrized description of the war of 1914-'18 as between Britain and Germany.

Constant	A	K	C	B	E
Unit	year^{-1}	year^{-1}	year^{-1}	year^{-1}	year^{-1}
Order of magnitude	1	$A + 0.3$	$1._6 \times 10^2$	1 $\geqslant 0.5$	10^{-2} 0.5×10^{-2}
Reference	(78)	(80)	(89)	(103) and (145)	(21) and (97)

Constant	$D + F$	K_{21}/K_{12}	C_{21}/C_{12}		E_{21}/E_{12}
Unit	year^{-1}	unity	unity		unity
Order of magnitude	5	N_1/N_2	$\sqrt{(N_1/N_2)}$		$(N_1/N_2)^{0.7}$
Reference	(112)	(84)	(159)		(22)

PART II B

GREGARIOUSNESS AND WAR-FEVER*

Introduction

The unity of a nation in war for defense or attack was attributed by Trotter (37), in a book which expresses British war-sentiments of the year 1917, to what he called the "instinct of the herd," a tendency to act together. On the contrary, the hypotheses (26) to (41) of the preceding theory of war-fever explain the observed national unity without making any reference to gregariousness. For example, the maximum value of η as shown in Fig. 3 for Case VI depends on C/B in the manner specified by (141) and (143). Here C is a constant expressing the reaction to hostile foreigners; and B is a constant expressing the wearying effect of war.

This diversity of explanation deserves further consideration.

It may be argued, against Trotter's thesis, that the mere fact of simultaneous and similar action by most of the persons in a nation is in itself no proof that their motive was gregarious. For example, most people sleep at night. But is it not probable that this custom arose primarily because each individual was similarly affected by darkness, and that any persuasion of one another to agree on a conventional bedtime was only a secondary modification? In the foregoing theory of war-fever, the observed national unity is attributed to the threat from the enemy acting simultaneously on all the citizens, just as the darkness of night acts on them all.

However we know from everyday observation how strong can be the control exerted by ideas of fashion, of good form, of doing the done thing, of conformity to the pattern of our culture. Prof. T. H. Pear (25) has recently discussed Dr. Ruth Benedict's statement that war itself is a social theme that may or may not be used in any culture. Pear concludes that "modern warfare is not due to simple instincts, nor is it inevitable," for it depends on culture-patterns which are not congenital, but are learned, and can be altered. The relations between what is congenital and what is learned appear to me personally as very important, but complicated, obscure, and controversial. For example the tendency to learn the national culture-pattern was regarded by Trotter as an instinct. I shall not here attempt to disentangle the inborn from the acquired, but only to describe how people conform to the example of their living fellows. That can be done with differential equations; whereas a discussion of conformity to the be-

*A brief account was read to the British Psychological Society at Brighton on April 11, 1942.

havior of past generations would require the more difficult technique of integral equations.

The effects of gregariousness are likely to show themselves before the outbreak of hostilities as a restraint exercised by the peace-loving majority on a fiery minority; whereas after the outbreak gregariousness instead constrains any who still feel pacific to conform to the war-fevered majority.

Pure fashion

To bring this gregarious effect prominently into consideration let us as a preliminary imagine the extreme and fanciful case in which the only motive is to agree with the majority. As in Case II we may put approximately

$$\xi + \eta = 1\,. \tag{160}$$

But now, to express the effect of conforming to the majority, $d\eta/dt$ has to be positive if $\eta > \xi$ and negative if $\eta < \xi$. The simplest formula of that type is $d\eta/dt = \eta - \xi$. But that would permit η to stray outside the possible range $0 \leqslant \eta \leqslant 1$. Buffers must therefore be inserted so as to make $d\eta/dt$ vanish at the termini, giving $d\eta/dt = \eta(\eta - \xi)\xi$. We must also insert a constant G in order that there may be the possibility of adjusting the speed of the theoretical changes to agree with observations. Thus we obtain

$$-d\xi/dt = d\eta/dt = G\eta(\eta - \xi)\xi\,, \tag{161}$$

in which G is the reciprocal of a time. That is the simplest formula which makes sense of pure gregariousness, and it is of the third degree in ξ and η jointly. The foregoing hypotheses (28) to (39) of the bipersonal theory are of the second degree in the variables jointly. In accordance with the habits of mathematicians, we have seen what can be done with terms of the second degree before proceeding to those of the third.

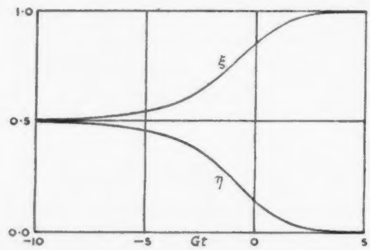
Although the present theory is mainly about multitudes, yet valuable hints may be gained by noticing the least number of people for which its concepts make sense. Two men alone on an island could quarrel and make peace; but neither of them could be said to be out of the island-fashion in clothes or speech. Thus the concept of fashion requires at least three people. In this sense the pair of equations (160) and (161) may be called *tripersonal.* The equations almost say so themselves; for, if $N = 2$, the only possibilities are $\eta = 1$, $\eta = \frac{1}{2}$, and $\eta = 0$, and in each of these three cases $\eta(\eta - \xi)\xi$ vanishes; whereas, if $N = 3$, then η can be $2/3$, making $\eta(\eta - \xi)\xi$ positive.

According to (161), the point $\xi = \eta$, at which there is no majority, cannot be crossed. The motion is away from this point on both sides of it. The existence of an uncrossable point is otherwise obvious. For, if the only motive in the population were to be fashionable, the fashion could never change. If peace and war were controlled by gregariousness alone, a nation would be either permanently at war or permanently at peace. Hence it is obvious that, in the real world, gregariousness acts, not as a prime cause of change, but as a modifier of changes caused by other motives. This accords, so far, with what Trotter wrote of effects as being *enhanced* by gregariousness.

The integral of (161) is

$$Gt + \text{Constant} = 2 \log(\eta - \tfrac{1}{2}) - \log \eta - \log(1 - \eta). \qquad (162)$$

A graph of this function is shown in Fig 4. There are two branches, according as η is greater or less than $\frac{1}{2}$. Change is slow when the population is either nearly unanimous or nearly equally divided. Change is most rapid when $\eta = 0.7887$ or 0.2113.



Gregariousness. Two folkways.

FIG. 4. Pure fashion. The abscissa is proportional to the time. The ordinate is the fraction of the population which follows one of two alternative folkways. The relation is theoretical, as specified by equations (160), (161), (162). The constant G is a measure of the intensity of gregariousness.

Case VIII. The outbreak of a symmetrical war, as modified by gregariousness

To examine the interaction between gregariousness and another strong motive, let us consider again the outbreak of a symmetrical war, which has been treated in Case II. But now let the bipersonal effect represented in Case II be combined with the gregarious effect represented in the section on pure fashion. An additive type of interaction will be assumed for mathematical convenience. Additive

interaction is also in accordance with a practice widely successful in theoretical physics. So let the new hypothesis be

$$d\eta/dt = C\eta\xi + G\eta(\eta - \xi)\xi\,. \tag{163}$$

Our other hypothesis is, as before,

$$\xi + \eta = 1\,. \tag{164}$$

Let

$$h = G/C\,, \tag{165}$$

so that h is a positive pure number expressing the strength of gregariousness relative to that of the motive for change. Accordingly (163) may be arranged as

$$\begin{aligned} d\eta/dt &= C\eta\xi\{1 + h(\eta - \xi)\} \\ &= C\eta(1-\eta)\{1 + h(2\eta - 1)\}\,. \end{aligned} \tag{166}$$

The chief question concerning (166) is whether $d\eta/dt$ can be negative; that is to say, whether gregariousness can prevent the outbreak of a war. The factors $C\eta(1 - \eta)$ are necessarily positive; and in the last factor

$$-1 \leqslant 2\eta - 1 \leqslant 1\,. \tag{167}$$

So $d\eta/dt$ can be negative if, and only if,

$$h > 1\,. \tag{168}$$

Let us consider separately the subcases $h > 1$, $h = 1$, and $0 \leqslant h < 1$.

Subcase $h > 1$. The last factor in (166) vanishes when

$$\eta = (h-1)/2h = \mathcal{E}\,, \text{ say,} \tag{169}$$

and $d\eta/dt \leqslant 0$ whenever $\eta < \mathcal{E}$; so that η, if initially near enough to zero, will, according to this theory, decrease to zero. We have already noted that gregariousness, if in sole control, would permit no deviation from a uniform and constant state of public opinion. We now see further in what circumstances gregariousness, although competing with a motive for change, may yet, according to this theory, convert the minds of a sufficiently small minority so as to make them agree with the majority. To express the same relations differently, let us suppose that gregariousness, as represented by G, is of fixed intensity while the motive for change, represented by C, has one or other of various constant values. If $C = 0$, the point $\eta = 1/2$ cannot be crossed. For larger constant values of C the uncrossable point is lower, being $\mathcal{E}$ as given by (169).

Subcase $h = 1$. Equation (166) becomes

$$d\eta/dt = C\eta(1-\eta)2\eta\,, \tag{170}$$

so that $d\eta/dt$ is positive for all values of η in the open interval $0 < \eta < 1$, but $d\eta/dt$ vanishes doubly at $\eta = 0$.

Subcase $0 \leqslant h < 1$. For a sufficiently large C, the uncrossable point lies outside the positive region of actual η; and then gregariousness modifies, but does not prevent, the conversion of persons from the mood {friendly, hostile} to the mood {hostile, friendly}. This happens, according to (169) when $h < 1$.

Integrals. Let us examine all three subcases more thoroughly by forming the integrals of (166) and (170). From (166), by separation of the variables and integration by partial fractions it may be shown, after a page or two of routine, that, if $h \neq 1$,

$$Ct + \text{Const.} = \frac{1}{1-h^2}\left\{\log\left(\frac{\eta}{1-\eta}\right) + h\log\left(\frac{\eta(1-\eta)}{4h^2(\eta-\mathcal{E})^2}\right)\right\}, \tag{171}$$

which can be verified more easily by differentiation. It is convenient to fix the arbitrary constant by the convention that

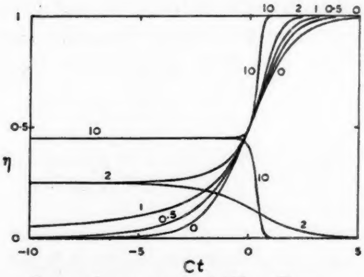
$$t = 0 \quad \text{when} \quad \eta = 1/2\,, \tag{172}$$

for with this choice the time is reckoned from about the beginning of hostilities. From (171) and (172) it follows that, if $h \neq 1$

$$\begin{aligned} Ct = \frac{1}{1-h^2}\,[\,(1+h)\log\eta + (h-1)\log(1-\eta) \\ -2h\{\log h + \log|\eta-\mathcal{E}|\}]\,. \end{aligned} \tag{173}$$

From this formula η has been computed as a function of Ct for various values of h, and is plotted in Figure 5.

The graph for $h = 0$ is of course that already mentioned in Case II. The graph for $h = 1/2$ resembles that for $h = 0$, but is steeper when $\eta > 1/2$ and less steep when $\eta < 1/2$. Leaving aside for the present the subcase $h = 1$, which requires a special formula, let us pass on to the graph for $h = 2$. It has two branches, which may be explained as follows. If by some disturbance, accidental in the sense that it is not taken into account in the present theory, η were momentarily raised to a value exceeding $\frac{1}{4}$ and there released, then the motives specified in the theory would carry η on towards unity in the manner shown by the upper branch of the curve. But if, on the contrary, the accidental disturbance raised η momentarily to a value less than $\frac{1}{4}$ and released it there, then η would sink back towards zero



Gregariousness and War Fever.

FIG. 5. Fashion competing with a motive for change.

Different theoretical ways in which a symmetrical war can begin, or can be avoided. The abscissa Ct is proportional to the time. The ordinate η is the fraction of the population that is eager for war. The numbers beside the curves are the values of h, which measures the intensity of gregariousness relative to that of threats.

The same diagram applies to the end of a symmetrical war with time running in the same sense. The ordinate is then ω_s, the fraction of the surviving population which is openly in favor of peace, as explained in Case IX.

in the manner shown by the lower branch of the curve. For $h = 10$ the curve shows a long interval of suspense followed by abrupt alternative decisions. The curves for $h = 2$ and 10 may serve as types for all those for which $h > 1$, on the understanding that the uncrossable value of η, given by (169), increases with h to the limit $\frac{1}{2}$ as $h \to \infty$.

The integral for the subcase $h = 1$ is best found from (170) because of the vanishing denominator in (171). The integral of (170), subject to the convention (172), can be shown to be

$$Ct = 1 + \tfrac{1}{2}\left\{\log\left(\frac{\eta}{1-\eta}\right) - \frac{1}{\eta}\right\}. \tag{174}$$

Moreover, (174) can be shown to agree with the limit of (171) as $h \to 1$ provided that $\eta \neq 0$. The curve for $h = 1$ separates the one-branched curves from the two-branched curves. When $h = 1$, an infinitesimal accidental disturbance of η from the value zero will start η on a drift towards unity. But for small values of η the drift is much slower for $h = 1$ than for $h = 1/2$.

Comparison of the theory with the outbreak of war in 1914. The diagram of observations (Figure 1) must first be artificially symmetrized by drawing a curve midway between those for Britain and Germany, and then this mid-curve must be compared with the theoretical Figure 5. It is obvious that the uncrossable point $\mathcal{E}$ was less than 1/20; and that therefore, by (169), $h < 10/9$. But it is not pos-

sible from those uncertain observations to prove that h exceeeded zero. This is a justification for believing that the bipersonal theory expresses the main effect and that gregariousness comes in only as a modification.

Case IX. The termination of a symmetrical war as modified by gregariousness

Towards the end of a war, gregarious effects may be expected to show themselves when ρ^* people are feeling tired but are concealing their weariness because they notice that others are patriotically doing the same. In a later phase some of these same people may notice others openly admitting that they have had enough of the war and may consequently do likewise.

This case may be regarded as a modification of Case IV. As in Case IV we assume that

$$\rho + \omega = 1 - \theta_T . \tag{175}$$

By a process of reasoning similar to that which led up to the hypothesis (163) let us now copy the bipersonal effect from Case IV and add a gregarious term from the theory of pure fashion, and so arrive at

$$d\omega/dt = (D + F)\rho\omega + G\omega(\omega - \rho)\rho . \tag{176}$$

If θ_T, which may be about 0.02, could be neglected, then (175) and (176) could be obtained from (164) and (163), respectively, by the simple substitution of ω for η, of ρ for ξ, and of $D + F$ for C. Consequently the same substitution would make the integrals of Case VIII applicable to Case IX, except for the arbitrary constant.

But, if it is desired to take θ_T into account, this can best be done by transforming the variables to fractions of the surviving population by putting

$$\rho/(1 - \theta_T) = \rho_s , \qquad \omega/(1 - \theta_T) = \omega_s . \tag{177}$$

For then (175) becomes

$$\rho_s + \omega_s = 1 , \tag{178}$$

and (176) can be arranged as

$$d\omega_s/dt = (1 - \theta_T)(D + F)\omega_s(1 - \omega_s)\{1 + h_s(2\omega_s - 1)\} , \tag{179}$$

in which

$$h_s = G(1 - \theta_T)/(D + F) ; \tag{180}$$

so that these equations are closely similar to those of Case VIII. It remains to fix the arbitrary constant. As in Case IV, we may rea-

sonably suppose that hostilities end at or about the time when $\omega_s = 1/2$, and so we can conveniently fix the arbitrary constant by the convention that

$$t = T \quad \text{when} \quad \omega_s = 1/2 . \tag{181}$$

This having been done, the integrals and the diagram of Case VIII become applicable to Case IX by the substitution of ω_s for η, of $(1 - \theta_T)(D + F)(t - T)$ for Ct, and of h_s for h. This useful substitution will be referred to as - - - - - - - - - - (182)

The most noteworthy effect of gregariousness, in Case IX, is that a considerable minority, openly in favor of peace, may remain almost constant for a long time prior to the end of the war, provided that $h_s > 1$. This behavior is in contrast with that in Case IV, where ω could be constant only when zero. There were in fact during 1916, 1917, and 1918 minorities opposed to the war in both Britain and Germany, as noted in Part I A. They were persistent, but whether they were constant is not clear.

MENTOR: In all the foregoing theory, you take the effect of the threat from the enemy-nation as acting always in the same sense. But in fact reversals have occurred. One might say that moderate threats weld a nation together; but very severe threats, following on military disaster, crack it in pieces. Witness the revolt of the Fascist Grand Council against Mussolini in July 1943, and the revolt of some German generals against Hitler in July 1944. How do you account for that on your theory?

AUTHOR: I have to leave it out as being a more complicated phenomenon. The last bit of theory is only about the change of mood from {hostile, weary} to {weary, hostile}. That change, to an open longing for peace, may have had a little to do with the revolts which you mention. But their main motive appears to have been discontent with the government for having failed to protect the nation. In the latter aspect they resembled the discontent in Britain with Mr. Neville Chamberlain in May, 1940; except that British parliamentary customs allowed his replacement to be effected in an orderly manner. The mood of "discontent with own government" is not on our present list. Theories, in their beginning, must omit.

Case X. The interaction of gregariousness with indignation in Britain during the Sudeten crisis of September, 1938

The reluctant acceptance in Britain of the Munich agreement of September 30, 1938, may be regarded as an instance in which gre-

gariousness prevailed over indignation. For the political fashion in Britain at that time was appeasement towards Germany and Italy. Yet the aggressive German behavior towards Czechoslovakia from early summer onwards aroused strong indignation in Britain; many Britishers wished somehow peacefully to quell Hitler by threats; and a minority, η_1, of Britishers would have preferred war with Germany to the cession of the Sudetenland, which was agreed to at Munich. These various attitudes can be seen in the Official Report (24) of the debates in the House of Commons on October 3, 4, 5, 6. It will suffice to refer to the speeches of Mr. Duff Cooper (Cols. 29 to 40), of Sir Samuel Hoare and Mr. Dalton (Cols. 150, 151) and to the motion (Cols. 557, 558) "That this House approves the policy of His Majesty's Government by which war was averted in the recent crisis and supports their efforts to secure a lasting peace"; this was carried by 366 to 144. We may conclude that, counting the tellers,

$$\eta_1 < 146/(146+368) = 0.28 \tag{183}$$

if the House of Commons represented the nation, which it seldom does precisely.

It is further evident from the history that, between September 1st and 28th,

$$d\eta_1/dt > 0 \,. \tag{184}$$

The order of magnitude of $d\eta_1/dt$ is not easy to estimate; perhaps, at a guess, η_1 may have risen from 0.01 to 0.10 in a month; so that

$d\eta_1/dt$ was of the order of unity during September. (185)

Many speakers in the aforesaid debate, while deploring the situation, expressed relief and satisfaction at the peaceful settlement. Chamberlain had described it as "peace for our time." From this it may be inferred that, just after the Munich agreement of September 30th,

$$d\eta_1/dt < 0 \,. \tag{186}$$

It further seems proper to assume that η_1, regarded as a function of time, was continuous across the crisis. - - - - (187)

Let suffix 1 indicate Britain and suffix 2 indicate Germany. Let us neglect β and ρ, so that, as in Cases II and VIII

$$\xi_1 + \eta_1 = 1 \,. \tag{188}$$

But the situation will not be assumed to be symmetrical.

The gregarious part of $d\eta_1/dt$ will be the same as in Case VIII, equation (163), namely $G\eta(\eta-\xi)\xi$, except that G, η, ξ will now become G_1, η_1, ξ_1. The international part of $d\eta_1/dt$ will be the unsymmet-

rical term $C_{12}\ \xi_1\ \eta_2$ quoted from (32). On adding these parts, we have

$$d\eta_1/dt = \xi_1\{C_{12}\eta_2 + G_1(2\eta_1 - 1)\eta_1\}\,. \tag{189}$$

Instead of writing an equation of motion for German moods, it will suffice to regard η_2 as a known external disturbance to British moods. The change in German η_2 can then be approximately represented by a sudden jump. Let η_2' be the value of η_2 just before the Munich agreement and η_2'' the value just after. It appeared to the British that

$$\eta_2' > \eta_2''\,. \tag{190}$$

On inserting in turn (184) and (186) in (189), we obtain:

just before the agreement

$$C_{12}\eta_2' + G_1(2\eta_1 - 1)\eta_1 > 0\ ; \tag{191}$$

just after the agreement

$$C_{12}\eta_2'' + G_1(2\eta_1 - 1)\eta_1 < 0\,. \tag{192}$$

From these two inequalities it follows, in view of (187), that $C_{12}(\eta_2' - \eta_2'') > 0$; and therefore, in view of (190), that

$$C_{12} > 0\,, \tag{193}$$

as was expected. Next, (183) implies that

$$1 - 2\eta_1 > 0\,, \tag{194}$$

and so (191) and (192) may be rearranged as

$$\frac{\eta_2''}{\eta_1} < \frac{G_1}{C_{12}}(1 - 2\eta_1) < \frac{\eta_2'}{\eta_1}. \tag{195}$$

We have thus a proof that $G_1 > 0$, and a hint as to how, with a little more information, the value of G_1 could be estimated. To make a guess, for illustration, suppose that $\eta_1 = 0.1$, $\eta_2' = 0.6$, $\eta_2'' = 0.2$; then would $2.5 < G_1/C_{12} < 7.5$. This range corresponds, as we should expect, to values of G/C which, in the symmetrical Case VIII, are large enough to give curves having two branches. The other branch, leading to war, was followed in August and September, 1939.

PART II C

Chemical Analogies

Much of the theory of Parts I B and II A was developed from the simple notion of the probability of a binary encounter, as introduced by Kermack and McKendrick (15) into their 1927 theory of epidem-

ics. Accordingly each of the terms in the second members of equations (26) to (39) is of the second degree in the variables. In chemical language the reactions would be bimolecular. But, as the reagents are persons, we had better call the reactions *bipersonal.* The simple approximation in the Cases II and IV leads to an integral commonly used by chemists; see (23, p. 643). Mole's (22) theory of the ignition of explosive gases has terms resembling, at least in sign, the author's linear theory of the instability of peace, as mentioned by Richardson (30). Those who are more familiar than the present author with the literature of physical chemistry may know where in it to find either helpful analogies or ready-made integrals of the complete set of differential equations (26) to (39), or of substantial parts thereof.

Dr. Ernest B. Ludlam, whose researches on the explosion of phosphorus vapor are mentioned in standard textbooks, has, at various times since 1936, drawn the author's attention to a social analogue to the activated molecules of chemistry and kindly permits it to be mentioned here. Common observation shows us that only a few of the persons who hold an opinion are sufficiently energetic to propagate it; just as only a few of the molecules, which have the formula suited to a reaction, are sufficiently energetic to react. Moreover, the unusual social energy abides at first in a small chain of leaders and their contacts; just as, in a chemical chain, the energy set free by the reaction is not immediately scattered among the whole vesselful of molecules. These suggestions, which may prove to be valuable, are mentioned here lest they be forgotten. But the author does not feel inclined to develop them at present. According to Semenoff (33), after van't Hoff in 1884 and Arrhenius in 1889 had stated the kinetics of bimolecular reactions, chemists pursued their researches for 24 years before Bodenstein in 1913 introduced the conception of chains of reactions. So a psychologist may excuse himself for taking time for deliberation. The author's theory is not yet at the stage of Arrhenius, for it takes no account of the different energy of different persons in the same dual mood.

REFERENCES

1. Bernhardi, F. von. Germany and the next war. Trans. by Powles, A. H. London: Edward Arnold, 1914.
2. British Association for the Advancement of Science. *Report of the Annual Meeting,* 1938, 277-334.
3. Buxton, Mrs. C. R. A weekly survey of the foreign press. *Cambridge Magazine,* Cambridge, England, 1915-1919.
4. Calvert, E. R. Capital punishment in the twentieth century. London: Putnams, 1930.
5. *Commonsense,* a London periodical current during the first world war.

6. Dumas, S. and Vedel-Petersen, K. O. Losses of life caused by war. Oxford: Clarendon Press, 1923.
7. Encyclopaedia Britannica. New York: Encyclopaedia Britannica, Inc., 1929, edition XIV.
8. German, a. J'Accuse, trans. by Gray, A. New York and London: Hodder and Stoughton, 1915.
9. *Glasgow Herald.* Daily Newspaper, Glasgow, Britain.
10. Graham, J. W. Conscription and conscience. London: Allen and Unwin, 1922.
11. Hirst, F. W. The consequences of the war to Great Britain. Oxford: Univ. Press, 1934.
12. Hitler, A. Mein Kampf, translated by Murphy, J. London: Hurst and Blackett, 1939.
13. James, W. The varieties of religious experience. London: Longmans Green, 1902.
14. *Keesing's contemporary archives,* weekly periodical. Bristol, England.
15. Kermack, W. O. and McKendrick, A. G. Mathematical theory of epidemics. *Proc. Roy. Soc. Lond.* A, 1927, 115, 700-722.
16. Kermack, W. O. and McKendrick, A. G. The solution of sets of simultaneous integral equations related to the equation of Volterra. *Proc. Lond. Math. Soc.,* 1936, 41, 462-482.
17. Kruger, P. Memoirs. London: Fisher Unwin, 1902.
18. Lloyd George, D. War memoirs. London: Odhams, 1938.
19. Matthew, Saint. Gospel, revised version. Oxford: Univ. Press, 1891.
20. McCarthy, J. A history of our own times. London: Chatto and Windus, 1892.
21. Mendelssohn Bartholdy, A. The war and German society. New Haven: Yale Univ. Press, 1937.
22. Mole, G. The ignition of explosive gases. *Proc. Phys. Soc. Lond.,* 1936, 48, 857-864.
23. Nernst, W. Theoretical chemistry. Trans. by Codd, L. W. London: Macmillan, 1923.
24. *Parliamentary debates, official report* (alias *Hansard*). London: H. M. Stationery Office.
25. Pear, T. H. Peace, war and culture patterns. *Bulletin of the John Rylands Library,* Manchester, 1948, 31, 3-30.
26. Plato. Timaeus.
27. Rashevsky, N. Mathematical theory of human relations. Bloomington, Indiana: Principia Press, 1947.
28. Richardson, L. F. Mathematical psychology of war. Oxford: Wm. Hunt, 1919. (A few copies remain with the author.)
29. Richardson, L. F. The behavior of an Osglim lamp. *Proc. Roy. Soc. Lond.* A, 1937, 163, 380-390.
30. Richardson, L. F. Generalized foreign politics. *British Journ. Psychol. Monog. Supplt.* No. 23, 1939.
31. Richardson, L. F. The number of nations on each side of a war. London: *Journ. Roy. Statistical Soc.,* 1946, 109, 130-156.

32. Richardson, L. F. Arms and insecurity, a microfilm, 35 mm. Kilmun, Argyll, Britain: from the author, $5, 1947.

33. Semenoff, N. Chemical kinetics and chain reactions. Oxford: Clarendon Press, 1935.

34. Soviet Union, official history of the communist party of the. Moscow: Foreign Languages Publishing House, 1939.

35. Spillane, J. P. In The neuroses in war, edited by Miller, E. London: Macmillan, 1940.

36. *Times, The.* Daily newspaper, London.

37. Trotter, W. Instincts of the herd in war and peace. London: Fisher Unwin, 1921.

38. *U. D. C.*, a periodical, London: Union of Democratic Control.

39. Vedel-Petersen, see Dumas.

40. Volterra, V. Theory of functionals, edited by Fantippié, trans. by Long, M. Glasgow: Blackie & Son, 1930.

41. Volterra, V. Leçons sur la théorie mathématique de la lutte pour la vie. Paris: Gauthier-Villars, 1931.

42. Wright, M. B. In The neuroses in war, edited by Miller, E. London: Macmillan, 1940.

PSYCHOMETRIKA—VOL. 13, NO. 4
DECEMBER, 1948

THE ROLE OF CORRELATION IN ANALYSIS OF VARIANCE

CLYDE H. COOMBS*
UNIVERSITY OF MICHIGAN

A test of the significance of a row or column agent in an analysis of variance may be expressed in the form of correlation between the agent and the variate. A test of the significance of interaction variance may be expressed in the form of correlation between the agents. These expressions are principally of theoretical interest in that the degree of significance in an F test or the value of a correlation coefficient may be controlled at will, or inadvertently, within certain limits.

In a recent article Peters (1) discussed the interpretation of interaction variance as correlation. There are a number of further interrelations to be pointed out and some interesting interpretations to be made. Peters' article just begins the study of the role of correlation in analysis of variance and does not emphasize some of the characteristics of both correlation and analysis of variance which are pointedly revealed by their joint study. The significance of this study lies not in the practical application of some of the formulas developed but rather in the design of research involving analysis of variance or correlational analysis.

I. *Effectiveness of an Agent and Intraclass Correlation*

Consider the simple case of an analysis of variance into three components. The levels of one agent are designated $1, 2, \cdots, i, \cdots, n$ and those of the other agent $1, 2, \cdots, j, \cdots, m$. With a single observation (or mean of observations) at each joint level ij, the observations may be represented by a table of n rows and m columns.

We shall proceed first to get an expression for the product-moment correlation between pairs of columns (or rows) [equation (10) or (16) and (17).] Let the mean of all observations be designated M and the score in any cell be designated X_{ij}.

$$M = \frac{\sum_i \sum_j X_{ij}}{n m}. \tag{1}$$

*This study was made possible by the Bureau of Psychological Services, Institute for Human Adjustment, Horace H. Rackham School of Graduate Studies, University of Michigan.

Designating the mean of any row as M_i, then

$$M_i = \frac{\sum_j X_{ij}}{m}. \tag{2}$$

The variance of a distribution of a sum of scores is

$$\sigma^2_{(X_{i1}+X_{i2}+\ldots+X_{im})} = \sigma^2_{X_{i1}} + \sigma^2_{X_{i2}} + \cdots + \sigma^2_{X_{im}} + 2r_{X_{i1}X_{i2}}\sigma_{X_{i1}}\sigma_{X_{i2}} + \cdots + 2r_{X_{i,m-1}X_{im}}\sigma_{X_{i,m-1}}\sigma_{X_{im}}. \tag{3}$$

Assuming homogeneity of variance from column to column and that the correlations between pairs of columns are equal, then

$$\sigma^2_{(X_{i1}+X_{i2}+\ldots+X_{im})} = m\,\sigma_j{}^2[1 + (m-1)r_j], \tag{4}$$

where $\sigma_j{}^2$ represents the variance of any column and r_j the product-moment correlation between any pair of columns.

The variance of the means of rows may be expressed as:

$$\sigma^2_{M_i} = \frac{1}{m^2}\sigma^2_{(X_{i1}+X_{i2}+\ldots+X_{im})} = \frac{\sigma_j{}^2}{m}[1 + (m-1)r_j]. \tag{5}$$

But

$$\sigma^2_{M_i} = \frac{\sum_i (M_i - M)^2}{n} \tag{6}$$

and

$$\sigma_j{}^2 = \frac{\sum_i (X_{ij} - M_j)^2}{n}. \tag{7}$$

But assuming homoscedasticity or homogeneity of variance, a better estimate of $\sigma_j{}^2$ may be obtained by combining the within columns sums of squares and dividing by the total number of cases. Hence, (7) becomes

$$\sigma_j{}^2 = \frac{\sum_j \sum_i (X_{ij} - M_j)^2}{n\,m}. \tag{8}$$

Substituting (6) and (8) in (5),

$$m^2 \sum_i (M_i - M)^2 = \sum_j \sum_i (X_{ij} - M_j)^2[1 + (m-1)r_j]. \tag{9}$$

Solving for r_j,

$$r_j = \frac{m^2 \sum_i (M_i - M)^2 - \sum_j \sum_i (X_{ij} - M_j)^2}{(m-1) \sum_j \sum_i (X_{ij} - M_j)^2}. \tag{10}$$

A single score in this design may be described alternatively as:

$$X_{ij} - M = (X_{ij} - M_j) + (M_j - M) \tag{11}$$

or

$$X_{ij} - M = (M_i - M) + (M_j - M) + d_{ij}. \tag{12}$$

Squaring and summing equation (11) over the rows and columns:

$$\sum_i \sum_j (X_{ij} - M)^2 = \sum_i \sum_j (X_{ij} - M_j)^2 + n \sum_j (M_j - M)^2. \tag{13}$$

Similarly for equation (12):

$$\sum_i \sum_j (X_{ij} - M)^2 = m \sum_i (M_i - M)^2 + n \sum_j (M_j - M)^2 + \sum_i \sum_j d_{ij}^2. \tag{14}$$

Subtracting equation (13) from (14) and rearranging terms:

$$\sum_i \sum_j (X_{ij} - M_j)^2 = m \sum_i (M_i - M)^2 + \sum_i \sum_j d_{ij}^2. \tag{15}$$

Substituting (15) in (10) and simplifying:

$$r_j = \frac{m(m-1) \sum_i (M_i - M)^2 - \sum_i \sum_j d_{ij}^2}{m(m-1) \sum_i (M_i - M)^2 + (m-1) \sum_i \sum_j d_{ij}^2}. \tag{16}$$

In a similar manner the equation for the product-moment correlation between rows (r_i) may be obtained and is as follows:

$$r_i = \frac{n(n-1) \sum_j (M_j - M)^2 - \sum_i \sum_j d_{ij}^2}{n(n-1) \sum_j (M_j - M)^2 + (n-1) \sum_i \sum_j d_{ij}^2}. \tag{17}$$

Our next step will be to express r_j and r_i in a form which will indicate their identity with Fisher's intraclass correlation [equations (24) and (25)], and then point out the well-known relation of F and intraclass correlation [equations (28) and (29).] Let us designate by V the mean sum of squares in analysis of variance. Then the mean sum of squares for rows is:

$$V_i = \frac{m \sum_i (M_i - M)^2}{n-1}. \tag{18}$$

Then

$$(n-1)(m-1)V_i = m(m-1)\sum_i (M_i - M)^2. \tag{19}$$

Similarly, the mean sum of squares for columns is:

$$V_j = \frac{n \sum_j (M_j - M)^2}{m-1} \tag{20}$$

and

$$(m-1)(n-1)\ V_j = n(n-1)\sum_j (M_j - M)^2, \tag{21}$$

and designating the error or remainder sum of squares V_e, then

$$V_e = \frac{\sum_i \sum_j d_{ij}^2}{(n-1)(m-1)} \tag{22}$$

and

$$(n-1)(m-1)V_e = \sum_i \sum_j d_{ij}^2. \tag{23}$$

Substituting (19) and (23) in (16):

$$r_j = \frac{(n-1)(m-1)V_i - (n-1)(m-1)V_e}{(n-1)(m-1)V_i + (n-1)(m-1)^2 V_e},$$

which simplifies to

$$r_j = \frac{V_i - V_e}{V_i + (m-1)V_e}. \tag{24}$$

Similarly, substituting (21) and (23) in (17) and simplifying, we have:

$$r_i = \frac{V_j - V_e}{V_j + (n-1)V_e}. \tag{25}$$

These last two equations will be immediately recognized as the same as the equation for Fisher's unbiased estimate of the intraclass correlation.

Solving equations (24) and (25) for V_e, we have, respectively,

$$V_e = \frac{V_i(1 - r_j)}{1 + (m-1)r_j}, \tag{26}$$

$$V_e = \frac{V_j(1-r_i)}{1+(n-1)r_i}. \tag{27}$$

From (26), rearranging terms

$$\frac{V_i}{V_e} = \frac{1+(m-1)r_j}{1-r_j} = F, \tag{28}$$

the F being Snedecor's F for testing the significance of the agent which varies from row to row.

Correspondingly, the test for the significance of the column agent is obtained from (27):

$$\frac{V_j}{V_e} = \frac{1+(n-1)r_i}{1-r_i} = F. \tag{29}$$

Hence, if

$$F = \frac{V_i}{V_e} \quad \text{or} \quad \frac{V_j}{V_e}$$

is found to be non-significant, then r_j or r_i, respectively, does not differ significantly from zero. If, on the other hand, the F is found to be significant, then it will sometimes be interesting and meaningful to express the relation as an intraclass correlation coefficient and thereby have an indication of the degree of the relationship.

Negative intraclass correlation will reveal itself in that V_i (or V_j) will be less than V_e. In a two-component analysis of variance with only one agent there is merely a V_B and a V_W representing, respectively, the estimate of universe variance with the agent varying and an estimate of universe variance with the agent fixed. If $V_B < V_W$, then the agent is generating negative intraclass correlation.

It should be further pointed out here that it is not the *interaction variance* which is being interpreted as correlation, but the effect of an agent on the variate is being described or measured by means of the correlation coefficient. As a matter of fact, in the three-component analysis discussed here, it is implicitly assumed that the interaction variance is an estimate of error variance only and, hence, that there is no correlation between the two agents. That interaction variance is a function of error and *correlation between agents* is the thesis of section IV of this paper.

An expression indicating the effect on residual variance of correlation between columns or correlation between rows may be readily obtained as follows:

Solving equation (16) for $m \sum_i (M_i - M)^2$, we have

$$m \sum_i (M_i - M)^2 = \frac{1 + (m-1) r_j}{(m-1)(1-r_j)} \sum_i \sum_j d^2, \tag{30}$$

and similarly from equation (17)

$$n \sum_j (M_j - M)^2 = \frac{1 + (n-1) r_i}{(n-1)(1-r_i)} \sum_i \sum_j d^2. \tag{31}$$

Substituting equation (30) and (31) in (14) and rearranging the terms, we have

$$\sum_i \sum_j d^2 = \frac{r_i r_j \sum_i \sum_j (X-M)^2}{(1-r_i) r_j + (1-r_j) r_i + (n-1)(1-r_i) r_i r_j + (m-1)(1-r_j) r_i r_j + r_i r_j}, \tag{32}$$

which, in the simplest case of two rows and two columns ($n = m = 2$), becomes

$$\sum_i \sum_j d^2 = \frac{(1-r_i)(1-r_j) \sum_i \sum_j (X-M)^2}{3 - r_i - r_j - r_i r_j}. \tag{33}$$

It is apparent that if $r_i = r_j = 0$, then one-third of the total sum of squares in a three-component analysis is residual and none of the F's would be significant.

II. *The* t-*test and Intraclass Correlation*

In those instances in which an agent is given only two values, the F test corresponds to the conventional application of the *t-test* to the significance of differences between two means.

Solving equation (28) for r_j in terms of F, we have:

$$r_j = \frac{F-1}{F+m-1}. \tag{34}$$

It has been shown (2) that

$$F = t^2 \tag{35}$$

in the case of two groups, and hence any t-test may be converted into an intraclass correlation.

Substituting (35) in (34),

$$r_i = \frac{t^2 - 1}{t^2 + m - 1}. \tag{36}$$

This formula, however, is not normally to be recommended unless the limitations of the correlation so obtained are very clearly understood and interpreted. These limitations are not peculiar to this correlation but pertain to those secured from formulas (24) and (25) or any other correlation including those obtained in the usual manner rather than via analysis of variance. Analysis of variance merely brings it home with more force. The value of the correlation between two variates is a function of the proportion of the variance of either associated with a common agent. The variance which an agent contributes to a variate is, of course, a direct function of its own variance. But the *proportion* of variance contributed by an agent is a function not only of the variance it contributes but a function of the variance being contributed by all other agents simultaneously. Hence, if the correlation in the universe is significantly non-zero, the actual value of the correlation secured on a sample can be manipulated within certain limits by controlling the variability of the agent in relation to the other agents in the universe.

This is of considerable significance to the use of correlation as a descriptive statistic and to the design of studies involving analysis of variance. In the opinion of the writer, in either of the above cases the degree of significance or relation of an agent to a variate is best tested or described when the agent in question and all other agents affecting the variate are permitted full normal variation.

III. ε^2 *and Intraclass Correlation*

In Section I the relation between F and intraclass correlation was investigated. The relation between ε^2 and F is known and given by the equation (2):

$$\varepsilon^2 = \frac{(K-1)\,F - (K-1)}{(K-1)\,F + N - K}. \tag{37}$$

From these two relations we are able to express the functional relation of ε^2 and the intraclass correlation. As a result of agent B we have an r_i given by (29). Substituting for F in (37), and rearranging terms:

$$\varepsilon^2 = \frac{n(m-1)\,r_i}{(m\,n-1) - (n-1)\,r_i}, \tag{38}$$

where the K of (37) is the same as m in our notation and the N of (37) is given by nm in our notation.

IV. *Application and Interpretation*

A recent study by Siegel and Stuckey (3) will be used here to show the application and interpretation of some of these formulas. Very briefly, this study consisted of observing the amount of water drunk and the amount of food eaten after each of four successive six-hour intervals for each of sixteen rats. Analyses of variance were made of the water intake and of the food intake. The results are reproduced in Tables I and 2 below.

TABLE 1
Water Intake

Source	Sum of Squares	df	Mean Square	F	Intraclass Correlations
time	968.00	3	322.67	76.3*	.825
animals	82.72	15	5.51	1.30	.07
residual	190.34	45	4.23		
total	1241.06	63			

*Significant at 1% level.

TABLE 2
Food Intake

Source	Sum of Squares	df	Mean Square	F	Intraclass Correlations
time	490.51	3	163.50	60.1*	.787
animals	13.94	15	0.93	2.92*	—.197
residual	122.40	45	2.72		
total	626.85	63			

*Significant at 1% level.

The last column of these tables contains the intraclass correlation obtained by use of equations (24) and (25). Let us consider the interpretation of the correlation .825 for water intake and the agent *time*. Here we have four classes or families, the observations on sixteen rats for each of four intervals of time. This high intraclass correlation signifies that the amount of water intake for the various members of a family tends to be much more similar than one would expect from chance. In other words, "time" is a significant agent in the amount of water intake—there are certain periods of the day or night when the animals are more inclined to drink than at other times.*

*It should be made clear that the interpretations being made here are literal interpretations of the data and of the results of the formulas applied. That these results might be explained in terms of controls present or absent in the conduct of this experiment is not the concern of the present writer.

Similarly, if we observe the high intraclass correlation of .787 for food intake with the agent *time* we may make a similar interpretation: that during certain intervals of the day and night the animals will be more or less likely to be eating.

Some interesting relations are also revealed by the intraclass r's for animals in these two studies. In the case of water intake, a non-significant correlation of +.07 is found for animals. Here a class or family consists of the four successive observations of a single animal's water intake. An intraclass r of zero indicates that the magnitude of one of the observations in a class has no relation to the magnitude of the others.

On the other hand, we find a significant negative intraclass correlation of —.20 for the food intake of an animal. This indicates that the four successive observations of the amount of food eaten are more *different* than would be expected by chance.

V. *Interaction and Correlation Between Agents*

Let us consider next the description of interaction in terms of correlation between the agents. Let us first consider a three-component analysis with two agents, A and B, with agent A partitioned over n rows and agent B partitioned over m columns with no replication.

A single score may then be represented as comprising the general mean plus a contribution from each of the agents and an error increment (ε), thus:

$$X = M + a + b + \varepsilon\,. \tag{39}$$

Hence,

$$x = X - M = a + b + \varepsilon\,. \tag{40}$$

Squaring, summing and dividing by the number of observations, we have

$$\sigma_x{}^2 = \sigma_a{}^2 + \sigma_b{}^2 + \sigma_\varepsilon{}^2 + 2\, r_{ab}\, \sigma_a\, \sigma_b\,, \tag{41}$$

assuming the errors to be uncorrelated with the agents A and B and letting r_{ab} be the correlation between agents.

In a four-component analysis there is replication of the design with, say, p observations in each group. Equation (41) is, then, the variance of the means of the mn groups. Multiplying (41) through by mnp we get the total sum of squares between groups, which may be written, in row and column notation, as

$$\begin{aligned} p\sum_j\sum_i(M_{ij}-M)^2 &= m\,p\sum_i(M_i-M)^2 + n\,p\sum_j(M_j-M)^2 \\ &+ m\,n\,p\,\sigma_\varepsilon{}^2 + 2\,p\,r_{ij}\Big[m\,n\sum_i(M_i-M)^2\sum_j(M_j-M)^2\Big]^{1/2}. \end{aligned} \tag{42}$$

The first two terms on the right-hand side of equation (42) when divided by the appropriate number of degrees of freedom give, respectively, the mean sum of squares for the row agent and the column agent (V_i and V_j). The third and fourth terms together give the interaction sum of squares, thus:

$$p \sum_j \sum_i d_{ij}^2 = m\,n\,p\,\sigma_\varepsilon^2 + 2\,p\,r_{ij}[m\,n \sum_i (M_i - M)^2 \sum_j (M_j - M)^2]^{1/2}. \tag{43}$$

The fourth component of the analysis is the independent estimate of sampling error coming from the within group sum of squares. Table 3 contains the analytical expressions for the various elements of a four-component analysis.

TABLE 3

Source	Σ of Squares	df	M Σ of Sq.
Rows	$m\,p \sum_i (M_i - M)^2$	$n-1$	$V_i = \dfrac{m\,p\,\Sigma(M_i - M)^2}{n-1}$
Cols.	$n\,p \sum_j (M_j - M)^2$	$m-1$	$V_j = \dfrac{n\,p\,\Sigma(M_j - M)^2}{m-1}$
ixj	$p \sum_i \sum_j d^2_{ij}$	$(n-1)(m-1)$	$V_I = \dfrac{p\,\Sigma\,\Sigma\,d^2}{(n-1)(m-1)}$
error	$\sum_i \sum_j \sum_h (X_{hij} - M_{ij})^2$	$nm(p-1)$	$V_\varepsilon = \dfrac{\Sigma\,\Sigma\,\Sigma(X_{hij} - M_{ij})^2}{n\,m(p-1)}$
Total	$\sum_j \sum_i \sum_h (X_{hij} - M)^2$	$nmp-1$	V_t

The first term on the right-hand side of equation (43), $mnp\sigma_\varepsilon^2$, is that part of the interaction sum of squares which is attributable to sampling error. In V_ε we have an independent estimate of the sampling variance of the universe. Hence, dividing $mnp\sigma_\varepsilon^2$ by its appropriate degrees of freedom and *assuming that these two estimates are equal,* we have

$$\frac{m\,n\,p\,\sigma_\varepsilon^2}{(m-1)(n-1)} = V_\varepsilon, \tag{44}$$

or

$$m\,n\,p\,\sigma_\varepsilon^2 = (m-1)(n-1)V_\varepsilon. \tag{45}$$

Also, as may be seen from Table 3,

$$p \sum_j \sum_i d_{ij}^2 = (n-1)(m-1) V_i , \tag{46}$$

$$m \sum_i (M_i - M)^2 = \frac{n-1}{p} V_i , \tag{47}$$

$$n \sum_j (M_j - M)^2 = \frac{m-1}{p} V_j . \tag{48}$$

Substituting from equations (45), (46), (47), and (48) into equation (43) we have

$$r_{ij} = \frac{V_I - V_\varepsilon}{2\sqrt{\dfrac{V_i V_j}{(n-1)(m-1)}}} . \tag{49}$$

It is obvious that if the interaction mean sum of squares is equal to the within group sum of squares, the correlation between agents is zero. It is apparent here, then, that the V_e of equation (22) in a three-component analysis is an estimate of error variance only when the agents are uncorrelated.

Dividing equation (49) through by V_ε and writing the F tests as

$$F_I = \frac{V_I}{V_\varepsilon}, \quad F_i = \frac{V_i}{V_\varepsilon}, \quad \text{and} \quad F_j = \frac{V_j}{V_\varepsilon},$$

(49) becomes

$$r_{ij} = \frac{F_I - 1}{2\sqrt{\dfrac{F_i F_j}{(n-1)(m-1)}}} . \tag{50}$$

It is apparent also from these equations that a mean interaction sum of squares less than the mean error sum of squares signifies a negative correlation between agents.

The generalization of these expressions to higher-order interactions should be pursued. It may be that higher-order interactions will require the postulation of a "mutual" correlation as a correlation between any number of agents.

REFERENCES

1. Peters, C. C. Interaction in analysis of variance interpreted as intercorrelation. *Psychol. Bull.*, 1944, 41, 287-99.
2. Peters, C. C. and Van Voorhis, W. R. *Statistical Procedures and Their Mathematical Bases.* New York: McGraw-Hill, Inc., 1940, p. 353.
3. Siegel, Paul S. and Stuckey, Helen L. The diurnal course of water and food intake in the normal mature rat. *J. comp. physiol. Psychol.*, 1947, 40, 365-70.

PSYCHOMETRIKA—VOL. 13, NO. 4
DECEMBER, 1948

A NOTE ON THE COMPUTATION OF A TABLE OF INTERCORRELATIONS

LEDYARD R TUCKER
EDUCATIONAL TESTING SERVICE

Outlined is the method used at present by the Educational Testing Service for computing intercorrelations from basic summations. This procedure is adapted to the use of high speed calculators in performing the calculations, and much of its value lies in the complete system of checks that is a part of the method. Besides the correlations that are the object of the procedure, covariances, means, standard deviations, and the number of cases are also recorded on the completed form to be available for further statistical steps.

In the June 1948 issue of *Psychometrika,* Kossack* noted that methods for calculating tables of intercorrelations from basic summations were not available in the literature. The purpose of this note is to present the method currently used at the Educational Testing Service. This method was developed, after careful consideration of several trial procedures, in order to minimize the labor of the calculations while maintaining computational accuracy and an adequate system of checks. Since the table of covariances between variables is often desired for further statistical steps, the operation of recording these values was incorporated into the method. High speed calculating machines are used in performing the computations.

Table 1 contains the calculations for a four-variable fictitious problem for which the computations are relatively simple, so that the various steps can be more readily followed. The body of the computing form in Table 1, which has a set of three rows and a column for each variable, is flanked at the left by two columns and at the top by three rows which are used to record reference values used in the calculations as well as the means and standard deviations. Three additional rows at the bottom are used for checking the calculations. The variable code numbers (Roman numerals in the example) are written in the row and column headed *Var.*

The notation used is as follows:

$i =$ individual,
$N =$ number of individuals,
$j =$ row variable,
$k =$ column variable,
$X_{ij} =$ score of individual i on variable j.

*Kossack, Carl F. On the computation of zero-order correlation coefficients. *Psychometrika,* 1948, 13, 91-93.

It is assumed that the sums of scores on the variables $(\sum_i X_{ij})$ and the sums of the products of scores for pairs of variables $(\sum_i X_{ij} X_{ik})$ have been obtained and can be recorded on the computing form. In Table 1 these values are printed in bold-face type. The sums are recorded both in the left column as $\sum X_j$ and in the top row as $\sum X_k$. The letter P is used to designate the sums of products so that

$$P_{jk} = \sum_i X_{ij} X_{ik} . \tag{1}$$

Note that the diagonal P's are the sums of squares.

A point of interest is that the Educational Testing Service has had special computing forms printed which are so designed that the International Business Machines tabulator can be made to print these summations directly on the computing forms, thus saving a transcription step.

When the $\sum X_j$'s, $\sum X_k$'s, and P's have been recorded, the sum of each array is obtained. The row of sums of scores is added and recorded at the right in the space $\sum_k (\sum X_k)$; the sum of the column of $\sum X_j$'s is recorded at the bottom in the space $\sum_j (\sum X_j)$. The sum of each column of P's is recorded in the row $\sum P$. These sums may be used to check the transcription of the values now recorded on the form.

The number of cases, N, is recorded in the upper left. N^2 is found and also recorded. This step should be carefully inspected for error, since it is not otherwise checked.

The next step is to compute the covariances, designated as C, by the formula:

$$C_{jk} = \frac{NP_{jk} - \sum X_j \sum X_k}{N^2} . \tag{2}$$

This formula was selected because it involves no derived values, (e.g., means), before the difference in the numerator is found. The rounding-off errors of derived values that are used in further computations often lead to serious errors in the resulting covariance. The computational procedure is to (a) multiply the P_{jk} by N, (b) subtract the product of the sums at the top of the column and the left of the row, and (c) divide by N^2. This computation for each C_{jk} can be accomplished on a calculating machine without recording intermediate values. The product dial of the machine in this case is not cleared. Care should be taken in setting up the decimal point on the multiplier and product dials so as to allow for the decimal digits of the results. The computation of the covariance for row variable II and column variable I in Table 1 is as follows:

$$C_{11.I} = \frac{152{,}100 \times 200 - 6{,}000 \times 5{,}000}{40{,}000},$$

$$= \frac{30{,}420{,}000 - 30{,}000{,}000}{40{,}000},$$

$$= \frac{420{,}000}{40{,}000},$$

$$= 10.5\,.$$

The calculating machine steps are given below.

Step	*Process*	*Keyboard*	*Multiplier Dial*	*Product Dial*
a)	Multiplication	152,100	200.0000	30,420,000.0000
b)	Cumulative negative multiplication	6,000	5,000.0000	420,000.0000
c)	Division	40,000	10.5000	0.0000

It is best to compute the C's in each column and to check these values before proceeding to the next column. The check for a column of C's is accomplished by the following formula:

$$\sum_j C_{jk} = \frac{N \sum_j P_{jk} - \sum_j (\sum X_j) \sum X_k}{N^2}. \tag{3}$$

The $\sum P$ for a column is used as if it were a P, and a C is computed as above using the $\sum X_k$ at the top of the column and $\sum_j (\sum X_j)$ recorded at the lower left. This C for $\sum P$ is recorded in the row *Check* $\sum C$. The sum of the C's in the column is found and recorded in the row $\sum C$. These last two values should check, allowing for rounding-off error. For example, the *Check* $\sum C$ for the first column is computed as follows:

$$(\text{Check } \sum C)_I = \frac{434{,}366 \times 200 - 6{,}000 \times 14{,}222}{40{,}000},$$

$$= \frac{86{,}873{,}200 - 85{,}332{,}000}{40{,}000};$$

$$= \frac{1{,}541{,}200}{40{,}000},$$

$$= 38.5300.$$

The sum of the C's in the first column is also 38.5300, which checks exactly. All four columns in the illustrative example check exactly; however, this is not necessary for completely accurate work. At times the column sums will differ from the check values by one or two points in the last digit. This is the rounding-off error mentioned above. When the first column of C's has been checked, the individual C entries for the column may be copied into the first row, thus saving time and labor. Similarly the second column of C's may be copied into the second row, etc. This step is permissible because each column is checked separately. Actually all off-diagonal C's are double checked.

Consider the diagonal C's. Here j and k are identical so that equations (1) and (2) become:

$$P_{kk} = \sum_i X_{ik}^2 ; \tag{4}$$

$$C_{kk} = \frac{NP_{kk} - (\sum_i X_{ik})^2}{N^2} ,$$

$$= \frac{\sum_i X_{ik}^2}{N} - (\sum_i X_{ik}/N)^2 . \tag{5}$$

Equation (5) is recognized as the formula for the variance of scores on variable k. Thus:

$$C_{kk} = \sigma_k^2 . \tag{6}$$

Accordingly, the square roots of the diagonal C's are found and are recorded in the row σ_k. For example, the diagonal C for variable I is 25.000 . Its square root is 5.00 , which is recorded in row σ_k and column I. Similarly the σ_k for variable II is 7.00 , which is the square root of 49.0000 , the diagonal C for variable II.

A preparatory step for the computation of the correlations, r's , is to record the reciprocals of the σ_k's in the column $1/\sigma_j$. For example, the reciprocal of 5.00, σ_I , is .200 , which is recorded in the $1/\sigma_j$ cell for variable I .

Equation (7) gives the relation between a correlation and the corresponding covariance:

$$r_{jk} = \frac{C_{jk}}{\sigma_j \sigma_k} . \tag{7}$$

In the present method the terms are regrouped so that:

$$r_{jk} = C_{jk} \frac{1}{\sigma_j} / \sigma_k . \tag{8}$$

Computationally this means that (a) one of the covariances, C_{jk} , is

multiplied by the $1/\sigma_j$ at the left of the row and (b) this product, which is left in the product dial, is divided by the σ_k at the top of the column. The result is recorded in the r row under the covariance. For example, for row variable I and column variable II:

$$
\begin{aligned}
r_{I.II} &= \frac{10.5000 \times .200}{7.00}, \\
&= \frac{2.1000}{7.00}, \\
&= .30 .
\end{aligned}
$$

The use of the reciprocals of the σ's makes it possible to perform the foregoing calculation as a continuous machine process without having to record intermediate values, e.g., the products of the pairs of σ's.

All values in the r rows are computed in order to obtain checks on the calculations. Ideally, the table of r's should be symmetrical, with unity in the diagonal cells. The rows of the computed r's should be compared with the corresponding columns in order that calculating errors may be detected. It is to be noted that symmetrical correlations are computed using different σ's and $1/\sigma$'s so that the two calculations do not repeat the identical numbers. Thus identical errors cannot be repeated. Sometimes the two symmetrical r's differ by one in the last digit due to rounding-off error in the $1/\sigma$'s. In these cases, the reciprocals can be carried to more places and the correct correlation computed.

The correlation of unity for the diagonals supplies a check on the computation of the σ's and the $1/\sigma$'s. At times when insufficient decimal places are used for σ and $1/\sigma$, a diagonal r may differ from unity by one in the last place. The diagonal r for variable IV is a case in point. Usually such a difference can be ignored.

A final step is to compute the means, which are recorded in the second row at the top. This is accomplished by dividing the sums in row $\sum X_k$ by N. These computations are checked by dividing $\sum_k \sum X_k$ by N, which yields $Ch. \sum M_k$. The means are summed to $\sum M_k$, which should agree with $Ch. \sum M_k$ within the fraction involved in rounding-off error. Thus the computing form contains the original summations, number of cases, means, standard deviations, covariances, and correlations. These can be used as a reference for various statistical steps that follow.

TABLE 1
Intercorrelation Computing Form

		Var.		I	II	III	IV	
$N = 200$ $N^2 = 40000$								
ΣX_j	$\frac{1}{\sigma_j}$	ΣX_k M_k σ_k		6000 30.00 5.00	5000 25.00 7.00	1170 5.85 1.75	2052 10.26 1.24	$\Sigma_k \Sigma X_k = 14222$ Ch. $\Sigma M_k = 71.11$ $\Sigma M_k = 71.11$
6000	.200	I	P C r	185000 25.0000 1.00	152100 10.5000 .30	35451 1.7550 .20	61815 1.2750 .21	
5000	.143	II	P C r	152100 10.5000 .30	134800 49.0000 1.00	29003 —1.2350 —.10	52076 3.8800 .45	
1170	.571	III	P C r	35451 1.7550 .20	29003 —1.2350 —.10	7457 3.0625 1.00	12158 .7690 .35	
2052	.806	IV	P C r	61815 1.2750 .21	52076 3.8800 .45	12158 .7690 .35	21359 1.5274 .99	
$\Sigma_j \Sigma X_j = 14222$		ΣP		434366	367979	84069	147408	
		Check ΣC		38.5300	62.1450	4.3515	7.4514	
		ΣC		38.5300	62.1450	4.3515	7.4514	

PSYCHOMETRIKA—VOL. 13, NO. 4
DECEMBER, 1948

FACTOR ANALYSIS I: SOME EFFECTS OF CHANCE ERROR

D. R. SAUNDERS*
UNIVERSITY OF ILLINOIS

Ignorance concerning the standard error of individual factor loadings and their differences has been a major obstacle to the proper interpretation of factorial results. The effects of three types of experimental error (selection of variables, selection of subjects and selection of scores) are here reported. In order to handle the errors of rotation systematically, it has been necessary to introduce a new semi-analytical criterion for the attainment of simple structure. Variability in results which may theoretically be eliminated is discussed under the heading of non-essential error.

Introduction

The results of a typical factor analysis rest on at least three steps, each involving experimental freedom. These are (a) the choosing of a number of types of measurement which can be made, (b) the selection or sampling of the population of subjects on which the measurements are to be made and (c) the actual measurement of the subjects with respect to the chosen variables. The resulting data are then turned over to the factor analyst, who applies two mathematical steps: (a) the computation of the matrix of simple correlations between every pair of variables and (b) the computation of factor loadings, or correlations of the variables with factors, according to some version of the equation.†

*The author wishes to acknowledge the kind counsel and encouragement received from Professor R. B. Cattell during the preparation of this manuscript.

†The following notational symbols will be used throughout this paper:

$a =$ a factor loading
$a' =$ loading without correction for attenuation
$h^2 =$ communality
$i =$ subscript denoting a variable of measurement
$j =$ subscript denoting a variable of measurement
$k =$ subscript denoting a factor
$n =$ total number of variables
$r_{ij} =$ correlation coefficient
$r'_{ij} =$ uncorrected correlation coefficient
$r_j =$ reliability coefficient
$w_{ik} =$ weight of variable i for factor k

$$a_{jk} = \frac{\sum_i w_{ik} r_{ij}}{\sqrt{\sum_j (w_{jk} \sum_i w_{ik} r_{ij})}}, \tag{1}$$

which has been derived algebraically by Guttman (2, Equation 25).

Corresponding to each of the experimental steps, both systematic and random errors may occur. These errors will affect the numerical results obtained by the factor analyst and lead him to inquire as to the probable error of any given factor loading. While the problem has thrice been attacked empirically, by Lorge & Morrison (4), by Mosier (6), and by McNemar (5), no results of general applicability were obtained. Indeed, the model adopted by Mosier fails to correspond to the usual first-order factor analysis, since it assumes that the error in each correlation coefficient is independent. Lorge and Morrison dealt with the restricted problem of (unrotated) principal components, while McNemar dealt only with unrotated centroid factors.

No systematic attack on the problem having been reported, it is the purpose of this paper to present a theoretical analysis of the probable error of any factor loading. The sources and types of error have already been indicated; these different *sources* of error operate independently, and may be most conveniently considered in the reverse order of their occurrence.

Errors of Measurement

Present-day factor analysis has had to incorporate the simplifying assumption of linearity. Since it is well known that a simple correlation coefficient is unaffected by constant errors, such errors may be dismissed from any discussion of the probable error of any loading computed according to equation (1).

However, it is also generally recognized that a correlation coefficient is attenuated by random errors of measurement and that corrections for attenuation may be made in terms of the reliability coefficients of the measurements involved in a given correlation. Thus (9)

$$r_{ij} = r'_{ij} / \sqrt{r_i r_j} \,. \tag{2}$$

---

w' = weight to be used with uncorrected correlations
H = number of variables whose vectors lie close to a hyperplane
K = number of factors extracted
N = number of subjects measured.
V = variance
σ = standard deviation

Combining equations 1 and 2, we find that

$$a_{jk}=\frac{1}{\sqrt{r_j}}\frac{\sum(w_{ik}/\sqrt{r_i})r'_{ij}}{\sqrt{\sum\left((w_{jk}/\sqrt{r_j})\sum(w_{ik}/\sqrt{r_i})r'_{ij}\right)}}, \tag{3}$$

whence

$$a_{jk}=\frac{1}{\sqrt{r_j}}\frac{\sum w'_{ik}r'_{ij}}{\sqrt{\sum w'_{jk}\sum w'_{ik}r'_{ij}}}, \tag{4}$$

or

$$a_{jk}=a'_{jk}/\sqrt{r_j}\,. \tag{5}$$

According to equation (5), the existence of simple structure is unaffected by attenuation [generalizing the result found by Spearman (10) for the situation involving only one general factor and clarifying the finding of Roff (7), which was stated only for centroid multiple-factor cases], but all the uncorrected loadings of an unreliable measurement are reduced systematically.

To obtain the best picture of the potentialities of different types of measurement, it is necessary to make the attenuation corrections. However, though they are unbiased, the corrections are themselves unreliable to an extent given by the formula

$$\sigma_{r_j}=\frac{1-r_j^2}{\sqrt{N}}. \tag{6}$$

By combining equations (5) and (6), together with the formula*

$$V_{a_{jk}}=\left(\frac{\partial a_{jk}}{\partial r_j}\right)^2 V_{r_j}\,, \tag{7}$$

we find that

$$\sigma_{a_{jk}}=\frac{(1-r_j^2)}{2\,r_j}\frac{a_{jk}}{\sqrt{N}}. \tag{8}$$

From equation (8) it can be seen that the percentage error in a_{jk} depends only on the reliability, r_j, and the number of cases used to evaluate it. The error tends to be larger for less reliable measurements. Since the chance errors of measurement will cause the same percentage error (in both sign and magnitude) in all the loadings of a given type of measurement, the effect of such errors may be ig-

*Equation (7) is the familiar propagation of error formula, given in this form by Shewhart (8). Its use here is equivalent to the procedure of Cureton (1) for estimating the error of a derived value.

nored when testing the significance of a difference between two such loadings.

Errors of Sampling

Thurstone (12) has discussed systematic errors of selection in order to show that, unless univariate selection is both composite and complete, simple structure is theoretically unaffected. However, the numerical values of the loadings are reduced in proportion to the amount of selection and the correlations between factors may be changed. At some point short of complete selection a factor will take on the familiar appearance of a "residual."

Thomson (11) has shown that even partial "multivariate selection" introduces normally unique factors into the common factor space and increases the rank of a correlation matrix. Thomson's multivariate selection, however, is not merely simultaneous univariate selection on more than one variable, as it involves disproportionate alteration of correlations. Simple multivariate selection, wherein the correlations of a type of measurement remain in proportion to its variability, does not increase the rank of a correlation matrix. Chance selection, as well as orthogonal multifactorial selection, is of this type.

Sampling error will further influence the variability of a sample of subjects in a given factor. If it is assumed that the distribution of the factor in the sampled population is normal, then the chance error in variability is given by the formula

$$\sigma_{\sigma_k} = \frac{\sigma_k}{\sqrt{2N}} . \tag{9}$$

Combining this with the known results of systematic selection leads to assignment of just this percentage error due to chance in all the loadings involving a given factor. When the significance of a difference between two loadings of the same factor is being tested, the effect of any possible selection errors, as such, may be ignored.

Errors of Rotation

Equation (1), by an appropriate choice of the weights, can be used to compute loadings in any factor, regardless of the actual methods of factoring or rotation employed. Therefore, granted the assumptions of factor analysis and ignoring errors of rotation, every loading is either zero or significantly different from zero, and there can be no such thing as a "residual." So-called residuals must represent either (a) factors with large errors of selection, or (b) factors which do not load any of the types of measurement heavily, or (c)

artifacts due to computational errors, including rounding, or (d) artifacts due to invalidity of the original assumptions. Thurstone's observation (13) that normalization of test scores prior to correlation tends to accentuate simple structure seems to indicate that (d) is an important source of residuals. However, when (c) has been effectively eliminated, both (a) and (b) should receive consideration.

Before the probable error of a loading due to chance errors of rotation can be evaluated, it is necessary to possess a suitable objective criterion of the attainment of the best simple structure. The criteria which have been proposed by Horst (3) and by Tucker (14) are sufficiently objective but do not readily lend themselves to the evaluation of the simple structure attained, at least in the form of a probable error of an individual loading. A new criterion is therefore proposed, which is similar to both of these, which may be utilized independently of any attenuation corrections, and which is specifically applicable to situations involving oblique or bipolar factors.

The hyperplanes of a factor structure necessarily pass through the origin of the factor space, and the deviation of a given test vector can be suitably measured as the angle it makes with the hyperplane at the origin. Goodness of fit can then be evaluated by summing the squares of the tangents of these angles, according to the formula

$$V_H = \frac{\sum^{H}\left(\frac{a^2}{h^2 - a^2}\right)^2}{H\{H - (K-1)\}}, \tag{10}$$

wherein, for a given position of the hyperplane, H is taken to include that group of test vectors lying nearest the hyperplane which makes the expression (for variance of estimate) a minimum. Other positions of the hyperplane may be sought which will reduce the expression still further. When V_H cannot be further reduced by any *small* displacement of the hyperplane, the criterion is satisfied. Other methods will probably be found more useful while the approximate positions of the hyperplanes are being discovered.

The variability in results due to chance error of rotation is now seen to stem from the first experimental step—the determination of the types of measurement to be applied to the subjects. The amount of error depends, for a given loading, on the communality of the measurement which is not explained by the loading, as given in the formula

$$\sigma_a = (h^2 - a^2)\sqrt{V_H}. \tag{11}$$

An equivalent statement is that the error is proportional to the length

of the projection of the test vector in the hyperplane. For purposes of significance testing, these errors must practically always be considered.

Systematic errors in the choice of variables will affect the assumptions underlying this treatment. It remains a problem for the individual researcher or critic to investigate whether such exist or are of any consequence.

Non-Essential Errors

There are three types of non-essential error which may affect the numerical results of a factor analysis. These are (a) missing measurements for some subjects on some variables, (b) rounding or other computational errors, and (c) inexact estimation of communalities prior to and during the analysis. All of these non-essential errors may theoretically be eliminated, but (a) may require a further comment.

Subjects for whom only incomplete data are available should probably be withheld entirely from the correlation matrix, in order to avoid increasing its rank through semi-independent errors in the correlation coefficients. However, data on such subjects may be advantageously employed when the reliability coefficients of the measurements actually included in the factorization are being estimated.

Summary

Those causes which may lead to non-ideal numerical results in factor analysis have been systematically discussed. Experimental error in selection or measurement of subjects causes only percentage errors in the results and often may be ignored for significance testing. Error in selection of the types of measurement to be made leads to imperfect rotation. To minimize errors of rotation, a new criterion of the best simple structure is proposed, which may be used to find the probable error of a single loading.

REFERENCES

1. Cureton, E. E. On certain estimated correlation functions and their standard errors. *J. exp. Educ.*, 1936, 4, 252-264.
2. Guttman, Louis. General theory and methods for matric factoring. *Psychometrika*, 1944, 9, 1-16.
3. Horst, Paul. A non-graphical method for transforming an arbitrary factor matrix into a simple structure factor matrix. *Psychometrika*, 1941, 6, 79-99.
4. Lorge, Irving and Morrison, N. The reliability of principal components. *Science*, 1938, 87, 491-492.
5. McNemar, Quinn. On the sampling errors of factor loadings. *Psychometrika*, 1941, 6, 141-152.

6. Mosier, Charles I. Influence of chance error on simple structure: an empirical investigation of the effect of chance error and estimated communalities on simple structure in factorial analysis. *Psychometrika*, 1939, 4, 33-44.
7. Roff, Merrill. The relation between results obtainable with raw and corrected correlation coefficients in multiple factor analysis. *Psychometrika*, 1937, 2, 35-39.
8. Shewhart, Walter A. Economic control of quality of manufactured product. New York: Van Nostrand, 1931.
9. Spearman, Charles. Demonstration of formulae for true measurement of correlation. *Amer. J. Psychol.*, 1907, 18, 161-169.
10. Spearman, Charles. The abilities of man. New York: Macmillan, 1927.
11. Thomson, Godfrey H. Factorial analysis of human ability. London: Univ. London Press, 1946.
12. Thurstone, L. L. The effects of selection in factor analysis. *Psychometrika*, 1945, 10, 165-198.
13. Thurstone, L. L. Multiple-factor analysis. Chicago: Univ. Chicago Press, 1947.
14. Tucker, Ledyard R. A semi-analytical method of factorial rotation to simple structure. *Psychometrika*, 1944, 9, 43-68.

r

t
h
t
a
p
t
A
b
e
o
g
h
ti
fo
a
g
w
g
re
tw
g
it
H
gr
of
th
th
an
or

PSYCHOMETRIKA—VOL. 13, NO. 4
DECEMBER. 1948

NOMOGRAM FOR THE TETRACHORIC CORRELATION COEFFICIENT

MAX HAMILTON
UNIVERSITY COLLEGE HOSPITAL
LONDON, ENGLAND

This article offers a new nomogram for the tetrachoric correlation coefficient, together with a correcting table. The development of the nomogram is described and directions for its use are included.

It is unnecessary here to enter into a discussion on the advantages and limitations of the tetrachoric correlation coefficient, as this has been frequently done before. In spite of the criticisms of the statisticians, it remains one of the most frequently used coefficients of association in psychological research. Because it is so difficult to compute, Karl Pearson published tables (5) to aid in its calculation, but they are little used because of the elaborate interpolation required. As the tables are unnecessarily accurate for practical purposes, a number of graphical aids have been designed to assist the computer. The earliest of these is the *abac* of Burt (1), which consisted of a family of curves on one diagram. They were actually computed for the colligation coefficient, but Vernon's (unpublished) Admiralty notes (6) have four of Burt's diagrams, computed for the tetrachoric correlation coefficient for dichotomies of 50, 34.5, 21.2 and 10%. These effectively cover the range, but interpolation between graphs is nearly always needed. The most popular of such aids are the computing diagrams of Thurstone *et al* (2). They are simple to read when working with data calculated to two decimal places, but this can occasionally give rise to some inaccuracy; working to three places requires two readings from each of two pairs of diagrams and interpolating between the pairs, a very time-consuming process. Other types of diagrams have been designed, although less well known, and each has its special advantages. One of the most recent is the set designed by Hayes (3). (It may be added that the description of the Hayes diagrams also includes a very good review of the uses, limitations, etc., of the tetrachoric correlation coefficient.) In general, it may be said that the two main disadvantages of most of the graphic methods are the need for interpolating arithmetically between pairs of diagrams, and sometimes the need to have some knowledge of the final result in order to know which diagrams to choose.

It was considered desirable to attempt the construction of a nomogram for the tetrachoric correlation coefficient which would have the advantage of being easy to read and of being contained in only one diagram. The true formula for the coefficient is impracticable for this purpose, but it was found that the empirical formula proposed by Karl Pearson himself was eminently satisfactory (4). This formula is rarely mentioned in current standard works on statistics applied to psychology, but it was used by Burt (1), who came to the conclusion that it was rather inaccurate, and particularly when the position of the dichotomy is in the tail of the normal distribution curve, beyond the 90%-10% point. The formula states that the tetrachoric correlation coefficient is obtained from the coefficient of colligation by multiplying the latter by ninety degrees and taking the sine of the angle so formed. With the four-fold table arranged as follows:

	−	+	
+	a	b	$a+b$
−	c	d	$c+d$
	$a+c$	$b+d$	N

Pearson's formula is therefore

$$r_{\text{tet.}} = \sin\left(\frac{\pi}{2}\cdot\frac{\sqrt{bc}-\sqrt{ad}}{\sqrt{bc}+\sqrt{ad}}\right), \tag{1}$$

where a, b, c, and d are the cell frequencies. This can be converted easily into

$$r_{\text{tet.}} = \cos\left(\pi\cdot\frac{\sqrt{ad}}{\sqrt{bc}+\sqrt{ad}}\right). \tag{2}$$

If r be read in degrees (which are re-scaled into cosines in the nomogram), the equation can be manipulated into

$$\sqrt{bc} = \sqrt{ad}\cdot\frac{(180-r^\circ)}{r^\circ}. \tag{3}$$

Although apparently suitable for the construction of a nomogram it was found in actual practice to be impossible. The reason for this is

that either in this form, or when the cell frequencies are converted into probabilities totalling unity, there are not four independent variables, but only three. The only way that could be found to deal with this problem was to divide the cell frequencies by one of them, e.g., d, when the equation becomes

$$\sqrt{BC} = \sqrt{A}\,\frac{(180 - r^\circ)}{r^\circ}\,, \tag{4}$$

where $A = a/d$, $B = b/d$, $C = c/d$ and $D = d/d = 1$.

Using the method of determinants, this equation can be used for the construction of a nomogram in two ways. Taking logs,

$$\log B + \log C = \log A + 2\log\frac{(180 - r^\circ)}{r^\circ}\,, \tag{5}$$

from which the following determinants are obtained:

$$\begin{vmatrix} -1 & \log B & 1 \\ 0 & \tfrac{1}{2}R & 1 \\ 1 & \log C & 1 \end{vmatrix} = 0 \quad\text{and}\quad \begin{vmatrix} -1 & \log \mathrm{A} & 1 \\ 0 & \tfrac{1}{2}R & 1 \\ 1 & 2.\log\dfrac{(180 - r)}{r} & 1 \end{vmatrix} = 0\,, \tag{6}$$

where R stands for a reference line. This nomogram consists of three parallel lines, the first being a logarithmic scale for B and C, the third a logarithmic scale of the same size for A and of twice the size for $\dfrac{(180-r)}{r}$, which is then rescaled for cos r, and the second a reference line half-way between the first and third.

A more compact nomogram in the shape of a circle can be constructed by converting the equation as it stands into the two determinants:

$$\begin{vmatrix} \dfrac{B}{B+1} & \dfrac{\sqrt{B}}{B+1} & 1 \\ \dfrac{1}{1+R^2} & \dfrac{-R}{1+R^2} & 1 \\ \dfrac{1}{1+\sqrt{C}} & 0 & 1 \end{vmatrix} = 0 \quad\text{and}\quad \begin{vmatrix} \dfrac{A}{A+1} & \dfrac{\sqrt{A}}{A+1} & 1 \\ \dfrac{1}{1+R^2} & \dfrac{-R}{1+R^2} & 1 \\ \dfrac{r}{180} & 0 & 1 \end{vmatrix} = 0\,. \tag{7}$$

The upper half of the circumference is scaled for A and B by locating the appropriate points for $B/(B+1)$ on the horizontal diameter and projecting on to the circumference. The horizontal is marked according to the formulas $1/(1+\sqrt{C})$ and $r/180$, and scaled for C and cos r. The lower half of the circumference forms the reference line. Of course, other varieties of nomogram can be designed, using this method of eliminating the superfluous variable, e.g., hexagonal and polar nomograms, but the two described are the simplest to use.

Preliminary trials soon showed that this had solved only half the problem. The nomogram tended to give results that were too high, but no regularity could be found in the discrepancies that were obtained. In the end, it was decided to attack the problem empirically, bearing in mind the supreme necessity for maintaining the simplicity of the nomogram. After trials of different methods, the following one was chosen: It is based on the fact that the formula gives a correct result when dichotomies of the data are at the medians, but one which is too big when the dichotomies move away from the medians. The percentage increase of the obtained result over the true was plotted against the dichotomy further from the midline, i.e., the more uneven dichotomy. In order to do so, it was necessary to find a large number of tetrachoric correlations known to be accurate. They were obtained by working backwards from Pearson's tables to obtain a a four-fold table for each coefficient. The dichotomies were chosen systematically to range from near the median to almost the 90-10% point. Thirty-six four-fold tables were obtained for each of the correlation coefficients selected: .10, .40, and .75, representing low, medium, and high correlations. In addition, twenty-one additional correlations were later calculated by interpolation from the tables, for the correlation of .75 at other dichotomies.

The curves, when plotted, are seen to ascend in a smooth fashion roughly in the shape of a parabola. The highest is the curve for $r = .10$, and the lowest that for $r = .75$. This arises from the nature of the curve, which plots the error as a percentage of the true figure. The curves broaden out from the origin because the dichotomies further from the medians allow of a greater variety of values, depending on the other dichotomy. The narrowest curve is for $r = .40$. Second-power parabolas were fitted to the curves, the equations for the fitted regression lines being:

$$\text{for } r = .10: \quad y = 1.851x^2 - 1.731x + 1.403; \tag{8}$$

$$\text{for } r = .40: \quad y = 1.351x^2 - 1.264x + 1.294; \tag{9}$$

$$\text{for } r = .75: \quad y = .830x^2 - .779x + 1.182; \tag{10}$$

where $y = \dfrac{\text{obtained } r}{\text{correct } r}$, and x is the area of the tail of the distribution curve, i.e., the smaller of the two parts of the dichotomy. It is difficult to give a measure of the accuracy of the fit of the regression lines, owing to the fan-like spreading of the curves. The *eta* coefficient is .922 and .982 for $r = .10$ and .40, respectively. On the other hand, the curvilinear correlation of the regression line (for $r = .40$) is .572. This rises to .780 when the data are suitably weighted to compensate for the extra data available at the distal end of the curve. As this was considered unsatisfactory, it was finally decided to use an empirical curve obtained graphically from the data, although eventually it appeared that the difference was completely negligible in the final nomogram, except for one point in each curve.

A new nomogram has to be obtained, based on the formula that the ratio of the obtained r to the correct r is given by the regression line, i.e.,

$$y = r_0/r_c\,, \quad \text{or} \quad y - r_0/r_c = 0\,. \tag{11}$$

From this, the determinant of the nomogram can be obtained:

$$\begin{vmatrix} 0 & r_0 & 1 \\ \dfrac{y}{y-1} & 0 & 1 \\ 1 & -r_c & 1 \end{vmatrix} = 0. \tag{12}$$

This nomogram consists of two parallel lines representing r_c and r_0, scaled in opposite directions, the distance between them being bridged by the line representing y, the correcting ratio. It may be added that the empirical regression lines and the calculated ones were significantly different only at the points representing the 90-10% dichotomies, where the difference came to 1 mm. in a nomogram 50 cms. in diameter. In effect, there are three nomograms, one for each of the three regression lines, having one line in common, the original scale for r. Since each nomogram covers only part of the range, the opportunity was taken to make separate magnified scales for ease in reading.

Finally, the results obtained by the nomogram, after correction, were compared with the true figures. It was found that, except for large correlations, the results were nearly always accurate to the second decimal place and were often within one or two units in the third decimal place. Where the dichotomy is no more uneven than 70-30%, the nomogram gives results which are accurate to within one or two digits of the third decimal place. It must be added that this finding is based on the figures for which the coefficient was calculated; for

others, the errors will tend to be larger, although owing to the method of construction, these are much smaller than may appear at first sight. Returning to the exception mentioned above, it was decided that it was necessary to introduce further corrections. A correcting table (Table 1) was prepared which takes into account the dichotomy so far ignored. The figures in this table are given signs which indicate their relation to the *absolute* value of the obtained correlation, i.e., the negative ones are subtracted and the positive ones added to the absolute value of the correlation, which is then given its appropriate sign.

TABLE 1
Correcting Table

The figures in these tables should be algebraically added to positive correlations and subtracted from negative ones.

S

		Other dichotomy	
		.15	.10
Most uneven dichotomy	.15	—.01	
	.10	—.01	—.01

M

		Other dichotomy					
		.45	.40	.30	.20	.15	.10
Most uneven dichotomy	.15			.01		—.01	
	.10	—.01		.01		—.01	—.02

L

		Other dichotomy						
		.50	.45	.40	.30	.20	.15	.10
Most uneven dichotomy	.40	—.01						
	.30	—.02	—.01					
	.20	—.03	—.02	—.01	.01	.01		
	.15	—.05	—.03	—.02		.01	.03	
	.10	—.05	—.04	—.02		.02	.04	.01

Description of the Nomogram

The nomogram consists of a circle and horizontal diameter, five short lines radiating from the centre, and three scales at the bottom,

from which the coefficient is read off. The upper half of the circumference carries a scale marked A, B, running from .05 to 20. The horizontal diameter carries a scale C on its upper side, marked 20 to .05. On its lower side, it is divided into regions marked S, M, and L, positive and negative, which are separated from each other by short vertical lines depending from the horizontal diameter. Radiating from the centre of the circle are five short lines designated according to the regions of the diameter. Each is scaled from .5 to .1, and it will be observed that the same point is marked .5 and .4. The scales at the bottom range from zero to .90, positive on the right side and negative on the left. The scale is actually extended to .95, but tetrachoric correlations of this size should not be taken seriously. To obtain a reading, three lines have to be drawn, the last of which runs from the horizontal diameter through a radiating line and ends in one of the scales at the bottom after crossing the vertical midline of the nomogram. The radiating line and the final scale are so chosen that they have the same letter and sign as the region of the horizontal diameter from which starts the third line drawn.

How to Use the Nomogram

The two variables to be correlated are arranged with the direction negative-to-positive (low-to-high, absent-to-present) running upwards and to the right. The frequencies are summed and entered into the outer cells.

	—	+	
+	a	b	$a+b$
—	c	d	$c+d$
	$a+c$	$b+d$	$a+b+c+d=N$

For the first step, the cell frequencies a, b, and c are divided by d, thus:

$a/d=A$	$b/d=B$
$c/d=C$	$d/d=D=1$

This can be done with one setting of a slide-rule.

B and C are located on the scales appropriately marked and joined by a line which is prolonged to the lower half of the circumference. A is then located on the top scale and joined to the lower end of the line just drawn. The region in which this second line cuts the horizontal diameter is noted, since it determines which of the five radiating lines and which of the final scales at the bottom shall be used for the drawing of the last line. For the next step, the dichotomy furthest from the median must be found. It is obtained by taking the *smallest* of the four sums in the outer cells, $a + b$, $c + d$, $a + c$, or $b + d$, and dividing it by N. Inspection of the nomogram will show that the quotient does not have to be very accurate. This point is then marked on the appropriate radiating line, joined to the point where the *second* line cuts the horizontal diameter, and prolonged down to the appropriate scale. (The first and second lines drawn must not be confused, and the beginner will find it best to number them.) The answer is read off and corrected, if need be, from the correcting table, which is self-explanatory.

The procedure above may be modified to make both the calculation and the drawing easier. Instead of dividing by d, the cell frequencies may be divided by a; D is located on the A, B scale. The result is the same. The scales B and C can be interchanged if it is more convenient for the drawing of the first line. On many occasions, it will be found that to divide the cell frequencies by b or c is easier arithmetically. In that case, the first line is drawn using A and D, and locating them on the upper circumference and horizontal diameter, it does not matter which. The second line is joined, in the upper scale, to the remaining number, B or C, whichever was *not* the divisor. The third line is drawn as already described, but the sign of the correlation read from the scale must be reversed. There is no point in repeating the process with a different divisor, since this will give the same result, provided the drawing has been done accurately. Slight inaccuracies in the first two lines make very little difference to the result, but a little care should be taken over the third line.

It is not claimed that the nomogram is inherently better than the other devices for obtaining $r_{\text{tet.}}$. They all have their individual advantages and disadvantages, but it can reasonably be asserted that the nomogram is the quickest and simplest.

REFERENCES

1. Burt, Cyril. Mental and scholastic tests. London: P. S. King & Son. 1921.
2. Cheshire L, Saffir, M., and Thurstone, L. L. Computing diagrams for the tetrachoric correlation coefficient. Chicago: Univ. of Chicago Bookstore. 1933.

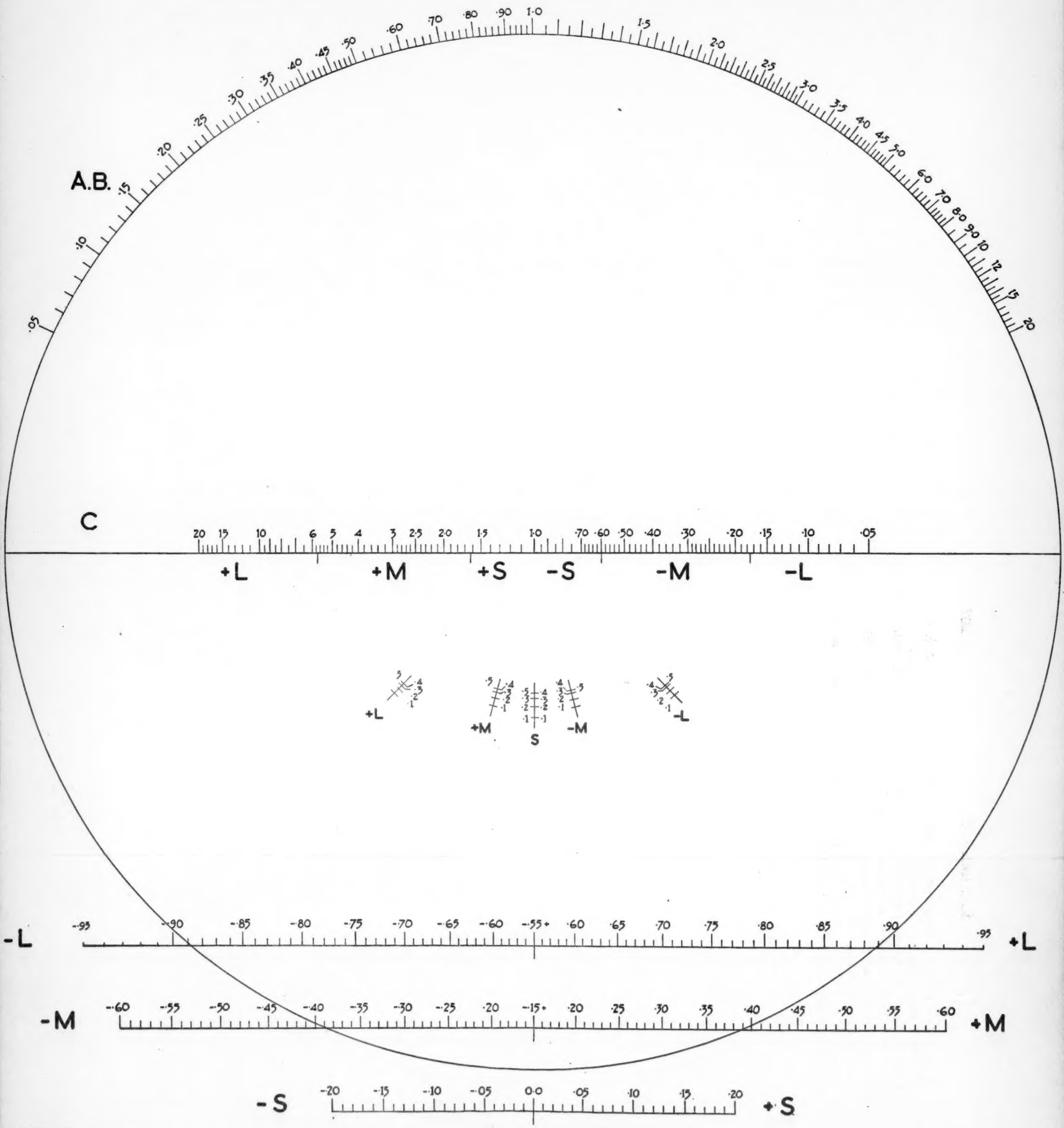
3. Hayes, Samuel P., Jr. Diagrams for computing tetrachoric correlation coefficients from percentage differences. *Psychometrika*, 1946, 11, 163-172.
4. Pearson, Karl. On the correlation of characters not quantitatively measurable. *Phil. Tran. Roy. Soc.* Series A, Vol. 195, 1901.
5. Pearson, Karl (Editor). Tables for statisticians and biometricians, Part II. Cambridge: Cambridge University Press. 1931.
6. Vernon, P. E. Notes on statistical methods in common use in vocational and educational research. 1941 (Unpublished).

NOMOGRAM FOR THE TETRACHORIC CORRELATION COEFFICIENT

DESIGNED BY MAX HAMILTON

NOMOGRAM FOR THE TETRACHORIC CORRELATION COEFFICIENT

DESIGNED BY MAX HAMILTON



P

of
w
O
te
to
co
re
re
co
te
va
sh

P
en
ti
fr
ar
es
ar
fo

PSYCHOMETRIKA—VOL. 13, NO. 4
DECEMBER, 1948

THE VALIDITY OF A PERSISTENCE TEST

JOHN W. FRENCH
EDUCATIONAL TESTING SERVICE

In order to raise the predictive efficiency of its college entrance test battery, the Educational Testing Service is working on the development of non-academic measures to supplement the standard aptitude and achievement examinations. A test of difficult number series problems was set up to measure persistence by tempting the students to give up early; the students were informed that some of the problems had no solution, and that full credit would be received by so marking them. This test was tried out and found to have some correlation with grades, while having no correlation with the other tests. Adding this test to the battery showed an appreciable rise in the battery's multiple correlation with grades.

As a part of its research program for increasing the efficiency of college entrance tests the Educational Testing Service has been working on the development of suitable tests of a non-academic sort. One such type of test comprises those which are called persistence tests, since they are designed to test persistence or something close to it. It was hoped that the efficiency of predicting college success could be increased by the use of such tests in conjunction with the regular academic tests, the purpose being to raise the multiple correlation of the test battery with a grade criterion above that which could be obtained by academic tests alone. A tendency for persistence tests to correlate to some extent with grades but to show only vanishing correlations with academic or intelligence tests has been shown by Howells (1) and by Ryans (2).

There has been some confusion of persistence with perseveration. Persistence (or sometimes "volitional perseveration") has most properly been understood to refer to action resulting from deliberate volition or will. Perseveration, on the other hand, properly refers to the frequently unfavorable tendency for one type of activity to continue after it has ceased to be called for. It tends to interfere with the establishment of a new activity. The distinction between persistence and perseveration has been made clear by Ryans (4), who has also found that there appears to be no relation between them.

The Test

Two investigators, Thornton (5) and Rethlingshafer (3), used

the centroid method of factor analysis with rotated axes to find "persistence" factors. They both distinguished a factor involving length of time spent on a test from a factor involving willingness to withstand discomfort. Thornton's best measure of the former was the length of time that students were willing to spend reading a story, where the lettering and punctuation became more and more confusing until reading became impossible. The test used in this study is probably a measure of this factor which was found by Thornton and called by him "keeping on at a task." It was designed to measure persistence by presenting a difficult task which unpersistent individuals would be tempted to give up.

The students taking the test were to remain unaware of its purpose. Accordingly, the test was given a false name suggested by its content material, number series problems. It was called the "Numerical Ingenuity Test." Directions and sample problems were given on one side of the test sheet. The test, constituting thirty number series problems, was presented on the other side. The Appendix shows the problems as they appeared. After finding the rule governing the construction of a series of six numbers, the students were to write down the seventh and eighth numbers in the series. The items ranged from medium difficulty to extremely difficult, with two items having no solution whatever (number 3 and number 8). In the directions the students were told that some of the items had no solution. They were to respond to such items by entering an X in the appropriate space. This feature was designed to encourage the unpersistent students to give up early in the hour, using a large number of X's.

This was a one-hour test. To keep track of the amount of time that the students worked, they were directed to write down a "time-number" whenever they wrote down or changed an answer, whether the answer was a number or an X. The time-number was that shown on a card at the front of the room. These were the numbers 1 through 60, numbering consecutively the minutes of the test hour and exhibited in sequence by the test supervisor

Scoring

The test was scored in six different experimental ways as follows (a high score in each case was supposed to be an indication of high persistence):

1. The highest time-number, a measure of the amount of time during which productive work was done. The range of this score was 13 to 60.
2. The number of reversals in the series of time-numbers, an indication of the amount of trying and retrying items pre-

viously attempted. Range: 0 to 14.

3. The number of items answered correctly. The range was 6 to 20 out of a possible 30. It can be seen from this that no student was so able at the task that he did not need to persist during the whole hour.
4. The number of items which were not answered by an *X*. Although this measure must be considerably affected by ability, it reflects the resistance of the student to giving up. Range: 9 to 29.
5. The number of time-numbers higher than 30, a measure of productivity during the second half-hour. It was thought that the second half-hour might discriminate between different amounts of persistence, while the first half-hour might not. Range: 0 to 14.
6. The time for going through the test once. This score was the highest time-number reached while taking the problems in numerical order and progressing forwards through the test. It was thought that a thorough attempt at each successive item before attempting the following item might denote a systematic persistent approach. Range: 5 to 52.

The Experimental Group

The Numerical Ingenuity Test was administered in August, 1946 to 345 highly motivated applicants to the Cooper Union Engineering School, Evening Division.* The results, however, are confined to the scores obtained on the 84 students who entered Cooper Union and received a full set of freshman grades.

Procedure

Included in the study were the variables listed in Table 1: the six persistence test scores (variables 1-6), five course grades (variables 7-12), a weighted average for freshman year (variable 13), and the scores received on ten academic entrance tests (variables 14-23) prepared by the Test Construction Department of the College Entrance Examination Board (now a part of the Educational Testing Service).

The product-moment intercorrelations of all variables were computed.† Table 1 gives the complete set of these intercorrelations. Using the values in this table, multiple correlations were computed for a

*The study was made possible by Professor Walter S. Watson, Admissions Officer of the Cooper Union, who made arrangements for the testing and acted as supervisor of this test.

†The computational work was carried out by the IBM installation of the Educational Testing Service, using a method devised by Ledyard R. Tucker.

TABLE 1
Table of Intercorrelations for all Variables
(decimal points omitted)
$N = 84$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1. Highest Time-number		52	19	37	70	19	21	08	14	28	27	12	30	09	01	15	08	16	20	–21	–07	06	–17
2. Reversals	52		05	32	52	–49	24	10	13	17	17	24	25	02	00	03	21	10	15	–07	–06	06	–09
3. Number Correct	19	05		49	10	03	16	09	16	20	–05	10	16	29	21	22	25	05	27	–10	–17	38	–28
4. Not Answered *X*	37	32	49		47	10	07	–08	05	16	–03	10	06	15	03	16	23	21	23	–11	03	22	–11
5. Time-numbers above 30	70	52	10	47		25	15	–07	11	19	09	05	14	–07	–08	–01	01	00	17	–31	–14	06	–25
6. First Time Through	19	–49	03	10	25		–23	–26	–08	–10	–19	–14	–24	–18	–19	–08	–14	–16	01	–28	–11	–11	–11
7. College Algebra	21	24	16	07	15	–23		63	33	08	56	16	78	60	63	43	04	–08	–18	14	10	00	–02
8. Descript. Geometry	08	10	09	–08	–07	–26	63		47	12	56	16	76	43	45	34	29	12	03	14	10	16	14
9. Engin. Drawing	14	13	16	05	11	–08	33	47		19	28	31	67	22	30	23	42	–07	01	–06	01	16	–08
10. English	28	17	20	16	19	–10	08	12	19		26	38	43	01	08	12	01	29	40	01	10	25	–03
11. Chemistry Test	27	17	–05	–03	09	–19	56	56	28	26		18	76	37	46	27	–02	29	13	34	27	12	26
12. Chemistry Lab.	12	24	10	10	05	–14	16	16	31	38	18		38	08	12	04	07	07	10	04	02	13	–06
13. Freshman Average	30	25	16	06	14	–24	78	76	67	43	76	38		50	57	42	21	12	07	16	16	20	08
14. SAT Math.*	09	02	29	15	–07	–18	60	43	22	01	37	08	50		77	58	18	07	–03	20	14	18	08
15. Math. Short Problems	01	00	21	03	–08	–19	63	45	30	08	46	12	57	77		65	07	04	–02	33	15	19	08
16. Math. Long Problems	15	03	22	16	–01	–08	43	34	23	12	27	04	42	58	65		17	04	02	20	17	27	16
17. Spatial Relations	08	21	25	23	01	–14	04	29	42	01	–02	07	21	18	07	17		09	09	–06	–11	18	–03
18. Reading	16	10	05	21	00	–16	–08	12	–07	29	29	07	12	07	04	04	09		51	44	46	52	41
19. Verbal	20	15	27	23	17	01	–18	03	01	40	13	10	07	–03	–02	02	09	51		09	18	48	27
20. Physics	–21	–07	–10	–11	–31	–28	14	14	–06	01	34	04	16	20	33	20	–06	44	09		53	24	56
21. Chemistry	–07	–06	–17	03	–14	–11	10	10	01	10	27	02	16	14	15	17	–11	46	18	53		26	50
22. Reasoning	06	06	38	22	06	–11	00	16	16	25	12	13	20	18	19	27	18	52	48	24	26		22
23. Science Background	–17	–09	–28	–11	–25	–11	–02	14	–08	–03	26	–06	08	08	08	16	–03	41	27	56	50	22	

*The mathematics section of the Scholastic Aptitude Test of the College Entrance Examination Board.

number of combinations of predictors, some of them including persistence tests and some not including persistence tests.

Results

Inspection of Table 1 will show that there is a definite tendency for the first two scores of the persistence test to correlate with grades and yet to have negligible or even negative correlations with the academic entrance tests. The other four scores of the test appear to be of little or no value.

As can be seen from the intercorrelations of the different persistence test scores, numbers 1, 2, and 4 appear to be measuring pretty much the same thing (presumably persistence), while the other three are unrelated. The negative correlations between persistence score no. 6, "First Time Through," and the academic variables probably mean that length of time spent on the test is an inverse measure of ability rather than a direct measure of persistence.

TABLE 2
Multiple Correlations Resulting from Combinations of Predictors

Criterion	Predictors	r's	Combinations	Multiple R's	Gain
Freshman Average	15. Math. Short Problems	.57	15,14,16	.58	
	14. SAT Math.	.50	15,1,2	.64	.07
	16. Math. Long Problems	.42	15,14,16,1,2	.65	.07
	1. Highest Time-number	.30			
	2. Reversals	.25			
College Algebra	15. Math. Short Problems	.63	15,2	.67	.04
	2. Reversals	.24			
English	19. Verbal	.40	19,18	.41	
	18. Reading	.29	19,1	.45	.05
	1. Highest Time-number	.28			
Chemistry Tests	15. Math. Short Problems	.46	15,18	.54	
	18. Reading	.29	15,1	.53	.07
	1. Highest Time-number	.27	15,18,1	.58	.04
Chemistry Laboratory	2. Reversals	.24	2,15	.27	.15
	15. Math. Short Problems.	.12			

Table 2 was prepared to show the amount by which the multiple correlations of test scores with grades were raised by the first two persistence test scores. The left-hand column of the table gives the grade criteria. The next column gives the number and name of the best academic predictors and the most successful persistence test

scores for the criterion named at the left. The correlation of each predictor with the criterion is given. The right-hand columns indicate combinations of test scores by number and their multiple correlations with the criterion named at the left. The last column shows the gain in r for which the persistence test is responsible.

A Caution

Since the number of cases used in this study is only 84, the correlations are subject to considerable error. The effect of such error can be especially dangerous, as it is here, where a large number of correlations are based on the available number of cases, and where interpretation is based on a relatively few good correlations selected from the many. The complete set of correlations is included in Table 1, so that the reader may judge for himself the extent of this effect.

Summary and Conclusions

A persistence test scored in six different ways was given to students who had also taken ten academic entrance tests and for whom course-grade data became available. Two of the persistence test scores showed appreciable correlations with grades, but negligible correlations with academic tests. This indicates that they were contributing new data toward grade prediction. The multiple correlation between predictors and freshman average was raised from .58 to .65 by the addition of the persistence test.

Because of the nature of these data, the results should not be regarded as entirely conclusive. They do, however, constitute encouragement toward the use of persistence tests in improving the predictive efficiency of college entrance test batteries.

REFERENCES

1. Howells, T. H. An experimental study of persistence. *J. abnorm. soc. Psychol.*, 1933, 28, 14-29.
2. Ryans, D. G. A study of the observed relationship between persistence test results, intelligence indices, and academic success. *J. educ. Psychol.*, 1938, 29, 573-580.
3. Rethlingshafer, D. The relation of tests of persistence to other measures of continuance of action. *J. abnorm. soc. Psychol.*, 1942, 37, 71-82.
4. Ryans, D. G. The measuring of persistence. *J. gen Psychol.*, 1938, 19, 79-96.
5. Thornton, G. R. A factor analysis of tests designed to measure persistence. *Psychol. Monog.*, 1939, 51, No. 229, p. 42.

APPENDIX

	1st	2nd	3rd	4th	5th	6th	7th	8th	indeterminate	number being displayed
1.	35	34	32	29	25	20	—	—	—	—
2.	3	7	11	15	19	23	—	—	—	—
3.	2	4	6	8	10	15	—	—	—	—
4.	13	14	15	13	17	12	—	—	—	—
5.	16	23	28	38	49	62	—	—	—	—
6.	140	139	137	131	118	95	—	—	—	—
7.	2	3	9	21	41	72	—	—	—	—
8.	1	2	4	8	17	34	—	—	—	—
9.	60	64	32	36	18	22	—	—	—	—
10.	51	47	44	43	46	57	—	—	—	—
11.	5	10	18	29	43	60	—	—	—	—
12.	100	95	85	80	70	65	—	—	—	—
13.	1	3	6	18	21	63	—	—	—	—
14.	11	14	10	15	9	16	—	—	—	—
15.	8	3	9	4	16	11	—	—	—	—
16.	4	8	11	22	25	50	—	—	—	—
17.	34	25	36	49	38	81	—	—	—	—
18.	13	15	51	53	35	37	—	—	—	—
19.	30	15	22	11	18	9	—	—	—	—
20.	137	73	41	25	17	13	—	—	—	—
21.	11	13	17	19	23	29	—	—	—	—
22.	59	10	46	21	37	28	—	—	—	—
23.	1	3	7	15	31	63	—	—	—	—
24.	25	82	58	16	91	49	—	—	—	—
25.	1	4	11	25	50	91	—	—	—	—
26.	1	7	3	9	5	1	—	—	—	—
27.	3	4	4	7	7	12	—	—	—	—
28.	1	3	9	19	35	61	—	—	—	—
29.	1	1	2	3	5	8	—	—	—	—
30.	13	16	22	24	28	36	—	—	—	—

PSYCHOMETRIKA—VOL. 13, NO. 4
DECEMBER, 1948

A WORKSHEET FOR TETRACHORIC r AND STANDARD ERROR OF TETRACHORIC r USING HAYES DIAGRAMS AND TABLES*

HOWARD W. GOHEEN AND SAMUEL KAVRUCK
VETERANS ADMINISTRATION

A worksheet simplifying the calculation of tetrachoric correlation coefficients and their standard errors is presented for use with Hayes' percentage difference method.

The development by Samuel P. Hayes, Jr. of computing diagrams based on percentage differences made available to research workers a short-cut method of approximating the value of the tetrachoric correlation coefficient (1). Tables for computing the standard error of the tetrachoric r were described by Hayes in a previous article (2). Where large numbers of the coefficients are to be computed and where an orderly and systematic arrangement of the data is desired for research, it has been found useful to have available a standardized worksheet, simple enough in construction for use by a statistical clerk. This paper presents such a worksheet which has been of value in the test construction section of a federal agency. The interpolation table in Hayes (1) is expanded, and a table facilitating the computation of the standard error of r_t is included.

TABLE A. INTERPOLATION OF r_t FROM DIAGRAMS (For use in step 6)

Ratio of % III : I	Roman numbers in table below correspond to diagram numbers in Hayes diagrams (1)			
	If + use r_t from diagram below		If − use r_t from diagram below	
1.000 : 1	I		I	
.861 : 1 .723 : 1 .584 : 1	I ± .25 difference I and II I ∓ .50 difference I and II I ± .75 difference I and II	+ if I < II − if I > II	I ± .25 difference I and VIII I ∓ .50 difference I and VIII I ± .75 difference I and VIII	+ if I < VIII − if I > VIII
.446 : 1	II		VIII	
.382 : 1 .317 : 1 .253 : 1	II ± .25 difference II and III II ∓ .50 difference II and III II ± .75 difference II and III	+ if II < III − if II > III	VIII ± .25 difference VIII and IX VIII ± .50 difference VIII and IX VIII ∓ .75 difference VIII and IX	+ if VIII < IX − if VIII > IX
.189 : 1	III		IX	
.156 : 1 .123 : 1 .091 : 1	III ± .25 difference III and IV III ∓ .50 difference III and IV III ∓ .75 difference III and IV	+ if III < IV − if III > IV	IX ± .25 difference IX and X IX ∓ .50 difference IX and X IX ± .75 difference IX and X	+ if IX < X − if IX > X
.058 : 1	IV		X	
.049 : 1 .040 : 1 .032 : 1	IV ± .25 difference IV and V IV ± .50 difference IV and V IV ∓ .75 difference IV and V	+ if IV < V − if IV > V	X ± .25 difference X and XI X ∓ .50 difference X and XI X ± .75 difference X and XI	+ if X < XI − if X > XI
.023 : 1	V		XI	
.005 : 1	VI		XII	

* We wish to thank Samuel P. Hayes, Jr. for his review and helpful suggestions leading to a simplification of the worksheet.

WORKSHEET FOR TETRACHORIC r AND STANDARD ERROR OF TETRACHORIC r
USING HAYES PERCENTAGE DIFFERENCE DIAGRAMS (1) AND TABLES (2)

Project: ______________________ N of sample: ______________

Variable I (X axis): ______________________ Point of Dichot: ______________

Variable II (Y axis): ______________________ Point of Dichot: .______________

Step 1. Set up 4-fold table.

	Above ____	Below ____	
Above ____	(a)	(b)	(a+b)
Below ____	(c)	(d)	(c+d)
	(a+c)	(b+d)	N=

Step 2. Determine small column and small row, then rearrange into 4-fold table below with small column at left, small row at top.

(a')	(b')	(a' + b')	← Small row
(c')	(d')	(c' + d')	
(a' + c') ↑ Small column	(b' + d')	N=	

Step 3. Compute % I: $\frac{a'}{a' + c'}$ = ____________ = .____________

Step 4. Compute % II: $\frac{b'}{b' + d'}$ = ____________ = .____________

% I is $\gtrless$ % II .·. $r_t \pm$ (circle what applies)

Step 5. Difference % I and % II: = .____________ (Locate on X axis in diagram)

Step 6. Compute % III: $\frac{a' + c'}{b' + d'}$ = ____________ = .____________ (Use diagram ____) Refer table A preceding page

Step 7. Compute % IV: $\frac{a' + b'}{N}$ = ____________ = .____________ (Locate on Y axis in diagram)

r_t = .____________

= $\boxed{}$ ← The sign of the r_t obtained must be examined in light of any changes made in setting up the 4-fold table in step 2. The sign of the r_t obtained must be reversed before entry here, if either the rows or columns, but not both, have been transposed in setting up the 4-fold table in step 2.

TABLE B. COMPUTATION OF σ_{r_t}

1. Compute % of combined groups above criterion: $\frac{a' + b'}{N}$ = ____________ = .____________
2. Compute % of combined groups falling in smaller group: $\frac{a' + c'}{N}$ = ____________ = .____________
3. r_t = .____________ N = ____________ $\sqrt{N}$: ____________
4. Refer to Hayes (2) table no. ________ (Interpolate if necessary)
5. $\frac{\sigma_{r_t} \sqrt{N}}{\sqrt{N}}$ = ____________ = (σ_{r_t})

REFERENCES

1. Hayes, Jr., S. P. Diagrams for computing tetrachoric correlation coefficients from percentage differences. *Psychometrika*, 1946, 11, 163-172.
2. ——— Tables of the standard error of tetrachoric correlation coefficients. *Psychometrika*, 1943, 8, 193-203.

PSYCHOMETRIKA—VOL. 13, NO. 4
DECEMBER, 1948

BOOK REVIEW

F. STUART CHAPIN. *Experimental Designs in Sociological Research.* New York: Harper and Brothers, 1947. Pp. 206.

In this brief compilation, Chapin demonstrates how certain types of "experimental designs" and statistical analyses apply to community investigations of the type that defy the usual experimental analysis, randomization, and control. In the seven chapters of the book, there is an attempt to present a systematic account of certain types of sociological research which the author identifies as experimental designs. One can of course quarrel with this extension of the use of this term beyond that to which we have become accustomed in the work of R. A. Fisher and his followers. Such objection, however, would lose sight of the important contributions offered by Chapin in pointing out certain efficient uses of inferential methods through selective control in the realistic situations of community living.

The objectives of this type of research are the noble ones of appraising by scientific methods the effectiveness of specific means of achieving certain socially desired ends and isolating cause-effect relationships in the complicated area of human relations. The nine research studies chosen to illustrate the particular experimental designs are representative of these large problems of social research. These studies are, however, explicitly characterized as being in a crude stage of development, as purely exploratory, and as leading to only tentative conclusions. Chapin points out that these applications of experimental designs may serve mainly as pre-tests of working hypotheses for future research and are necessarily subject to verification by replication. Unfortunately, it appears that these limiting characteristics will appy to almost all the research performed according to the models described in Chapin's book.

Within the first chapter on "natural social experiments by trial and error" is a brief but excellent account of the theory and practice of the experimental method. The acceptable criterion of experimentation is the actuality of human interference with the conditions that determine the phenomena under observation. The difficulties of this type of intervention, especially in real community situations, are indicated, and the remainder of the book is devoted to description and evaluation of feasible and socially acceptable techniques by means of which hypotheses about social phenomena may be tested objectively and statistically. The techniques are second best because of the practical and also theoretical difficulties in this type of research. However, the author does a good service by demonstrating how our second best can be made by proper uses of matching, selective control, and fractionation within the accumulated data. In the attainment of the goal of precision (within the limits available), a premium is placed on measurement, which is stressed in proper critical form throughout the book. Chapter 6 offers an account of the problems of reliable and valid measurement of social behavior under the title "Sociometric Scales Available for Control and the Measurement of Effects."

Chapin describes only three experimenal designs, although other types are alluded to. These three types are: (1) the cross-sectional design, which makes

controlled comparisons for a single datum; (2) a projected design or "before and after" study, in which we measure the effects of a social program or social force at some future date; and (3) an *ex post facto* design, in which some present effect is traced backward to an assumed causal complex of factors or forces at a prior date. Illustrations of each type of investigation are described in detail with an account of the variables studied, controls instituted, and discussion and interpretation of the results.

Cross-sectional design is illustrated by two examples, the essentials of which involve the selection of two groups which have been differentiated by having or not having had a particular experience or "influence" in the past. (It should be pointed out here that this differentiation was made in the past, and so the criterion of experiment by intervention is not in fact adhered to by Chapin.) This experience differential is the independent variable. The available members of the two groups are then systematically matched on a number of pertinent variables and attributes, either by individuals or in terms of the frequency distributions. For those remaining in the two matched groups, determinations are made of the differences in the chosen dependent variable or variables, and the appropriate critical ratios or multiple critical ratio (Guttman's method) are computed and interpreted. The hypotheses of Chapin's two illustrative examples are that there is a relationship between the duration of Boy Scout tenure in a particular area and subsequent participation in community activities and community adjustments of these Scouts four years after leaving the organization (Mandel, 1938); also, that there is a difference between groups who have had work relief and direct relief in their comparative effects on morale, adjustment, social participation, and social status (Chapin and John, 1939). Designs for these studies are presented in terms of flow-charts, and there is a detailed discussion of the results in terms of the factors of selection and the several matching controls instituted.

Four examples are given to illustrate what Chapin calls "projected experimental design." This is the classical pattern of "before and after" experiment that operates by pre-test and post-test, before and after the introduction of some condition or influence on one group and not on another equivalent group. The investigations cited to illustrate this design are the study of Dodd on the effects of a rural hygiene program on a community in Syria, Chapin's study on the various social effects of public housing in Minneapolis, Hill's study of the influence of staff stimulation on social participation and social adjustment among university students, and the investigation by Shulman on the effects of a certain type of controlled activity program on delinquency behavior in children. In these studies, control groups were matched with experimental groups before the introduction of the independent variable, and the final results of measurement are fractionated in several suggestive ways. Again the data are subjected to analyses of critical ratios.

Ex post facto design is illustrated by three examples, two in which self-comparison is made of the same individuals and families at present date and at a past date, and one in which comparinson is made of populations *in the same area* at present date and at some past date. One study by Barer (1945), was on delinquency before and after admission to a New Haven housing development. This is such a simple method and one which yields such sparse and uncontrolled data that successive replications should have been indicated as the only really acceptable data from such a design. The study by Christiansen (1938) on the effects of length of high-school education on economic adjustment allowed illustration of better controls by having the groups involved matched (in frequency

distributions) on the following six factors: age, sex, high-school marks, nationality of parents, father's occupation, and neighborhood status. It was carefully pointed out that the establishment of these six controls shrank the original sample of 2127 students down to a total of 290, a decline of more than 86%. About 98% were lost when individual matching was performed on the control factors. A study by Chapin *et al* (1946) on rentals and tuberculosis death rates was used as an illustration of the *ex post facto* design in which relations are studied on the same areas rather than the same individuals. This study offered more scope for discussing statistical criteria in the proper choice of factors to be controlled.

Chapin's book has several good features and several bad ones. It points out some of the techniques by means of which we can tentatively test hypotheses in real situations which are not ordinarily sufficiently under the investigator's control for experimentation in the classical fashion. To that end the author has rendered a distinct service. On the other hand, the title of the book and the handling of the material in it might lead the reader to believe that Chapin covers all the major techniques for this type of study. This is far from being true. There is no apparent appreciation here of the power of statistical inference beyond the critical ratio and the correlation coefficient, and the real assumptions underlying these two statistics are not fully considered in the free manner in which they are employed. Several quantitatively minded sociologists will be bothered by Chapin's relatively complete reliance on these methods. The question of the nature of the populations for which the resulting generalizations are made continues to be a large question, due to the large number of restrictive controls that are introduced in the matching procedures advocated. Chapin points out that scientific generalizations are to be stated with the qualification, "everything else being equal." What he tries to show is that we can approach this condition of having all the important things equal through selection and matching controls. This would tend to approach the laboratory situation, a desirable model. Great care must be exercised, however, in the wide and general application of results from the laboratory when it is not general laws but socially applicable facts that we are seeking.

Chapin's book will be read by large numbers of people and with varied reactions. This reviewer would predict that considerable good will be done by the volume at least in creating some provocation to apply research methods and more advanced statistical analyses to the important problems of social phenomena.

Department of the Army T. G. ANDREWS

JOHN C. FLANAGAN (Editor). *The Aviation Psychology Program in the Army Air Forces.* Army Air Forces Aviation Psychology Program Research Reports, Report No. 1. Washington, D. C.: U. S. Government Printing Office, 1948. Pp. xii + 316.

During the recent world war, an Aviation Psychology Program was established within the United States Army Air Forces to conduct research in the following major fields: (1) development and refinement of procedures for the initial selection and classification for individual specialties on the basis of aptitude, interests, personality, and experience; (2) advanced selection and classification for individual, crew, and unit assignments on the basis of aptitude, interests, personality, measured proficiency, and experience; (3) design of equipment, including cockpits, controls, instruments, and display systems with reference to the

human capacities of the personnel who operate this equipment; (4) training problems, including job analyses, evaluation of training standards, objective study of teaching methods, improvement of curricula and instructional materials, effectiveness of training aids and devices, selection and training of instructors and their evaluation, and optimal methods of retaining proficiency after training; (5) development of refinement of methods of measuring proficiency at various levels of training in the different types of officer and enlisted personnel duties; (6) psychological problems involved in the operation of new types of aircraft and weapons; (7) opinions, attitudes, and motivation of individuals and groups and the effects of propaganda on these; (8) problems of personnel management, leadership, morale, and personal adjustment; and (9) follow-up of personnel through training and subsequent AAF careers to determine what psychological measures or other records are effective in predicting later performance.

The results of this tremendous program of research are reported in nineteen volumes under the title of *Army Air Forces Aviation Psychology Program Research Reports*. The first of these reports is the one being reviewed here. The author states that the purpose of this book is to provide background information concerning the development of the entire research program and a brief summary of the more important contributions to the solution of Air-Force problems and the major implications for future work. The book is divided into three main parts. Part one is devoted to a general history of the planning and development of the program. Part two presents briefly the important research findings and accomplishments of the program with respect to specific Army Air Force problems during the war. Part three is devoted to a discussion of the more general contributions to psychological theory resulting from this research.

Since the book provides a history of this important integrated research program as well as a brief outline of work in different areas reported more fully in subsequent reports, it will constitute a valuable guide and reference work for psychologists in every field, especially for those who do not have sufficient time or reason for reading the entire nineteen reports.

It seems worthy of comment that in spite of the mass of heterogeneous data, directives, and findings with which the author has been forced to cope in writing this report, he has succeeded in making his book interesting to read as well as authoritative.

University of Southern California ANDREW L. COMREY

FREDERICK B. DAVIS (Editor). *The AAF Qualifying Examination.* Army Air Forces Aviation Psychology Program Research Reports, Report No. 6. Washington, D. C.: Government Printing Office, 1947. Pp. xvii + 266.

This book is one of the nineteen volumes which report the results of the tremendous research program carried on during the past war by the Army Air Forces Aviation Psychology Program. The section of this research program described here is that concerned with the development and evaluation of the AAF Qualifying Examination which was designed to eliminate applicants for flying training who were not sufficiently literate to become officers and who did not possess the requisite skills and aptitudes to successfuly complete flying training. In view of its purpose, the examination has properly been characterized by the author of this book as the most important single examination used by the Army Air Forces during World War II. The examination was designed primarily as a

power test for which seventeen different forms were developed, comprising a total of more than 2,900 items. A three-hour time limit was attached to the examination for administrative convenience.

The problems and techniques involved in the selection of appropriate kinds of items for the examination are discussed in conjunction with a short but excellent chapter devoted to such problems in test theory as item construction, item difficulty, item validity, intercorrelation of items, test purity, and test reliability. Many technical problems of statistics in relation to test construction and evaluation are discussed throughout the book. These discussions include some modifications of older techniques and some new methods such as one for determining the extent to which tests selected for their usefulness in conjunction with an established composite aptitude rating will increase the correlation between the aptitude rating and the criterion.

A chapter is devoted to each of several types of tests developed for use in connection with the Qualifying Examination. These include verbal tests, information tests, tests of practical judgment and reasoning, mechanical comprehension tests, and perceptual tests. Sample items are given along with data concerning the effectiveness with which each type of test was employed. In an additional chapter, certain miscellaneous types of items which were not used extensively in the Qualifying Examination are presented. These items concerned mathematics, interpretation of data, psychomotor abilities, and flexibility of attention.

In addition to the Qualifying Examination, four other examinations are briefly presented in one chapter. These included a test to select the top twenty per cent of cadets for exemption from further academic training in the pre-flight period, an examination to divide pilot graduates into second lieutenants and flight officers, a test of English expression for selecting likely instructors, and a test of aeronautics aptitude for secondary-school pupils.

The discussion of factor techniques is almost entirely limited to Kelley's principal-axis method, which was applied to certain tests of practical judgment, reasoning, and mechanical comprehension. The author fails to indicate the limitations of this method and the advantages of others for the purpose of discovering the underlying psychological variables involved in test performances.

Psychometricians will be interested in the handling of important statistical problems encountered in developing this examination as well as in the information reported concerning many important mental skills. Psychologists faced with problems of test construction for selection purposes will find much useful information in this book.

University of Southern California ANDREW L. COMREY

Revue de l'Institut International de Statistique, 1945, 13 and 1946, 14. The Hague, Netherlands: W. P. Van Stockum et Fils.

The field of the Revue, sponsored by the International Institute of Statistics, is statistics in a broad meaning of the term, from mathematical aspects to census figures and economic reports. Statistics as applied to psychological problems was not covered up through 1946, although there are a few references to educational journals. Because of conditions caused by the war, Volumes 13 and 14 appeared as single issues rather than as quarterly numbers.

Readers of *Psychometrika* will be chiefly interested in two phases of the

publication: a well-arranged annual bibliography of statistical articles in the principal languages and an occasional original article in the field of mathematical statistics. Major articles in French carry an English summary, and vice versa.

That the English-speaking countries are at present the most active in the development of statistics as a science is perhaps indicated by a count of the theoretical articles listed in the bibliographies. In the 1945 bibliography, 201 of the theoretical articles had appeared in English as contrasted with 41 in other languages; in the 1946 listing, 101 references were in English as opposed to 30 in other languages.

Washington University, St. Louis PHILIP H. DUBOIS

PSYCHOMETRIC SOCIETY

STATEMENT OF RECEIPTS AND DISBURSEMENTS FOR FISCAL YEAR ENDED JUNE 30, 1948

RECEIPTS:

Dues

Membership Year	Members		Student Members	
	No.	Amt. Pd.	No.	Amt. Pd.
1950	1	$ 5.00		
1949	4	20.00		
1948	256	1,280.00	23	$69.00
1947	44	217.00	5	9.00
1946	13	65.00		
1945	6	30.00		
1944	4	20.00		
1943	4	20.00		
1942	3	15.00		
1941	2	10.00		
Total	337	$1,682.00	26	$78.00

Total Dues Received	$1,760.00	
Miscellaneous Receipts	1.68	
Total Receipts		$1,761.68
DISBURSEMENTS		
Psychomet. Corp. (Single issue order with Society dues)	$ 1.25	
James L. Longley (Deficit, 1947 banquet)	21.76	
Bank Chg. (Print special check book)	3.25	
G. D. Bradshaw (return of overpayment)	.25	
Total Disbursements		26.51
Excess of Receipts over Disbursements		$1,735.17
Balance on hand, July 1, 1947		687.93
Balance on Hand, June 30, 1948		$2,423.10
Amount due Psychometric Corporation (90% of dues)		1,584.00
Adjusted Balance as of June 30, 1948		839.10
Bank Reconciliation as of June 30, 1948		
Balance as per bank statement		2,423.35
Check outstanding		.25
Balance on hand, June 30, 1948		$2,423.10

COMMENTS

The Society took in $1,761.68 this year, as compared with $1,290.25 last year, and its membership increased from 220 to 256.

After allowing for the amount due to the Psychometric Corporation, the adjusted excess of receipts over disbursements is $151.17, as compared with $130.59 last year.

Disbursements, other than the 90 per cent of membership dues paid to the Psychometric Corporation, remained nominal: $26.51 this year as compared with $14.86 last year.

PSYCHOMETRIC CORPORATION

STATEMENT OF RECEIPTS AND DISBURSEMENTS FOR FISCAL YEAR ENDED JUNE 30, 1948

RECEIPTS:			
Subscriptions, 1947			
Institutional (25¼) - - - - - - - - -	$252.50		
Individual (2) - - - - - - - - - - - -	10.00		
Total (27¼) - - - - - - - - - - - -		$ 262.50	
Subscriptions, 1948 (205) - - - - - - -		2,050.00	
Subscriptions, 1949 (3) - - - - - - - -		30.00	
Subscriptions, 1950 (1) - - - - - - - -		10.00	
Total Subscriptions (234¼) - - - - -		$2,352.50	
Back Volumes and Issues - - - - - - - -		874.71	
Total Subscriptions and Sales, at List Prices		$3,227.21	
Less discounts* and collection charges - -		237.94	
Net Receipts from Subscriptions and Sales			$ 2,989.27
Other Receipts			
Psychometric Monographs - - - - -		9.05	
Overpayments less underpayments - - -		31.90	
Total - - - - - - - - - - - - - - -			40.95
Total Receipts - - - - - - - - - - - -			$ 3,030.22

*Ten per cent discounts to subscription agencies; seventy per cent discounts to foreign subscribers in war-devastated areas.

DISBURSEMENTS:

Secretarial services			
Bertha Cable	$ 16.30		
Jessie Garcia	9.36		
Frank Kellmyer	3.80		
Dorothy Comer	166.25		
Total		$ 195.71	
Editorial services		300.00	
Stationery and postage		98.56	
Printing		1,577.92	
Bank charges		5.62	
Auditing and revising accounting system		400.00	
American Bonding Company		25.00	
Post Office box rent		3.00	
Overpayments returned		33.50	
Subscriptions cancelled		13.50	
Total Disbursements			$ 2,652.81
Excess of Receipts over Disbursements			$ 377.41
Balance on hand, July 1, 1947			8,144.46
Balance on Hand, June 30, 1948			$ 8,521.87
Amount due from Psychometric Society			1,584.00
Adjusted Balance as of June 30, 1948			$10,105.87

BANK RECONCILIATION AS OF JUNE 30, 1948

Balance as per bank statement	$8,692.52
Checks outstanding (2)	170.65
Balance on Hand, June 30, 1948	$8,521.87

COMMENTS

The Corporation took in $4,614.22 (counting the sum due from the Psychometric Society) this year, as compared with $3,341.64 last year. Disbursements were also higher: $2,652.81 as compared with $1,765.49. However, the adjusted balance of $10,105.87 is still $1,961.41 higher than last year's balance, which was $8,144.46.

There were 236¼ subscriptions in 1947-48, as compared with 251 in 1946-47. The fractional subscriptions represent changes whereby all subscriptions now start with the March issue of Psychometrika.

A
A
I
C
C
C
C
C
C
D
D
D
D
D
E
E

INDEX FOR VOLUME 13

AUTHORS